

RAFAEL LUÍS SILVA DIAS

**DESENVOLVIMENTO DE SISTEMA DE MONITORAMENTO DA
CONCENTRAÇÃO DE SÓLIDOS SUSPENSOS TOTAIS EM RESERVATÓRIOS
COM BASE EM SENSOR MULTIESPECTRAL ACOPLADO A AERONAVE
REMOTAMENTE PILOTADA E APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

Orientador: Demetrius David da Silva

Coorientadores: Cibele Hummel Amaral
Elpídio I. Fernandes Filho

VIÇOSA – MINAS GERAIS

2020

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

D541d
2020
Dias, Rafael Luís Silva, 1993-
Desenvolvimento de sistema de monitoramento da
concentração de sólidos suspensos totais em reservatórios com
base em sensor multiespectral acoplado a aeronave remotamente
pilotada e aprendizado de máquina / Rafael Luís Silva Dias. –
Viçosa, MG, 2020.
74 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Demetrius David da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.62-70.

1. Sensoriamento remoto. 2. Espectrometria. 3. Água -
Qualidade. 4. Monitoramento. 5. Imagens multiespectrais.
I. Universidade Federal de Viçosa. Departamento de Engenharia
Agrícola. Programa de Pós-Graduação em Engenharia Agrícola.
II. Título.

CDD 22. ed. 621.3678

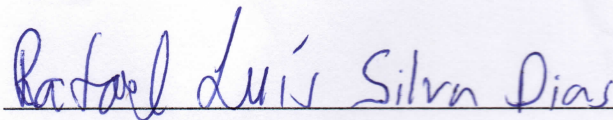
RAFAEL LUÍS SILVA DIAS

**DESENVOLVIMENTO DE SISTEMA DE MONITORAMENTO DA
CONCENTRAÇÃO DE SÓLIDOS SUSPENSOS TOTAIS EM RESERVATÓRIOS
COM BASE EM SENSOR MULTIESPECTRAL ACOPLADO A AERONAVE
REMOTAMENTE PILOTADA E APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

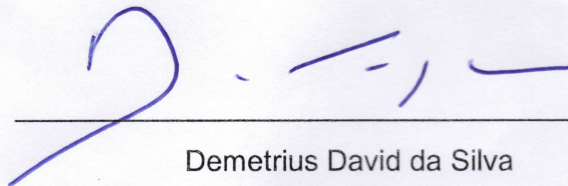
APROVADA: 22 de julho de 2020.

Assentimento:



Rafael Luís Silva Dias

Autor



Demetrius David da Silva

Orientador

*A Deus e a minha família, **dedico.***
*Sívio Bueno Pereira, **em memória.***

AGRADECIMENTOS

Primeiramente agradeço a Deus por sempre estar ao meu lado durante a minha caminhada, protegendo e dando alento nos momentos mais críticos. E pela intercessão de Nossa Senhora em minha vida.

Aos meus pais Walmir e Maria, minha fonte de amor, carinho e que nunca pouparam esforços e sempre abriram mão de muitos dos seus prazeres para garantir, a mim e meu irmão, uma boa formação. A meu irmão Bruno, por ser meu parceiro número um, me apoiando incontestemente sempre.

A Fernanda, pelo amor, companheirismo e amizade, que desde o início não poupou esforços para me incentivar a encarar este desafio e, durante todo esse tempo, mesmo de longe, foi e continua sendo meu porto seguro.

A meu orientador Demetrius, pelos conhecimentos e exemplo de profissionalismo a mim repassados, pela confiança e apoio durante a orientação. Aos meus coorientadores Elpídio e Cibele por me auxiliarem nos momentos de dificuldade, sem vocês não teria conseguido.

Aos meus estagiários Juliana e Thiago que foram fundamentais para a execução deste trabalho.

A toda equipe multidisciplinar que auxiliou em cada uma das campanhas de campo, em especial ao Ian e Laura do Departamento de Engenharia Civil, Jorge e Marcelo do Departamento de Engenharia Agrícola, Gustavo do Departamento de Solos e ao Corpo de Bombeiros Civis do Campus UFV.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos. *

Aos meus amigos e colegas do Centro Referência em Recursos Hídricos (CRRH) pela amizade, companheirismo e por compartilharem bons momentos e as trocas de conhecimentos em nosso dia a dia. Em especial ao irmão de orientação Erli Pinto pela caminhada conjunta e sempre nos apoiando.

Enfim, o meu muito obrigado a todos que me ajudaram e apoiaram de alguma forma durante o mestrado.

* O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

*“Por mais que as coisas avancem,
o básico, sempre será o
fundamental.”*

(Desconhecido)

RESUMO

DIAS, Rafael Luís Silva, M.Sc., Universidade Federal de Viçosa, julho de 2020. **Desenvolvimento de sistema de monitoramento da concentração de sólidos suspensos totais em reservatórios com base em sensor multiespectral acoplado a aeronave remotamente pilotada e aprendizado de máquina.** Orientador: Demetrius David da Silva. Coorientadores: Cibele Hummel Amaral e Elpídio Inácio Fernandes Filho.

O monitoramento da qualidade da água é fundamental para a manutenção e conservação dos recursos hídricos. O método convencional de monitoramento é feito por amostragens pontuais, que se limitam a representar a variabilidade espacial dos seus parâmetros, sendo expressivamente oneroso. Uma alternativa promissora constitui-se nas técnicas de sensoriamento remoto com o uso de sensores orbitais, que vêm sendo aplicadas para mapear parâmetros de qualidade da água, principalmente os Componentes Opticamente Ativos (COAs), como os Sólidos Suspensos Totais (SST), Clorofila-a (Chl-a) e Matéria Orgânica Dissolvida Colorida (CDOM). No entanto, o monitoramento de pequenos reservatórios com sensores orbitais sempre apresenta limitações, como as resoluções espectrais, espaciais ou temporais. Além disso, todos os sensores ópticos orbitais estão sujeitos a perturbações atmosféricas, além das nuvens, principalmente no período de chuva, justamente quando o carreamento de sedimentos é mais intenso. Nesse contexto, o objetivo deste trabalho foi desenvolver um sistema de monitoramento da concentração de Sólidos Suspensos Totais (SST) em reservatórios com base em sensores multiespectrais acoplados a uma Aeronave Remotamente Pilotada (RPA). Para isso, foi utilizado um sensor multiespectral *MicaSense RedEdge*, que possui cinco bandas nas faixas dos espectros do azul, verde, vermelho, *rededge* e infravermelho (NIR). Foram realizadas quatro campanhas de *in loco*, em diferentes estações do ano, variando desde o período de estiagem de 2019 ao período com condições de chuva extremas de 2020. Dois reservatórios de acumulação de água totalmente inseridos no bioma de Mata Atlântica foram escolhidos. Os dados foram submetidos a cinco diferentes algoritmos de aprendizado de máquina: *Random Forest* (RF), *Support Vector Machines Radial Sigma* (SVM-RS), *Multivariate Adaptive Regression Spline* (MARS), *Multiple Linear Regression* (MLR) e Cubist, a fim de estimar a concentração de SST nos reservatórios, tendo sido usados os índices

estatísticos raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE) e coeficiente de determinação (r^2) na comparação entre os valores preditos e observados. Os resultados mostraram que o *Recursive Feature Elimination* (RFE) reduziu significativamente o número de covariáveis dos modelos de aprendizado de máquina. O modelo SVM-RS foi o que teve o melhor ajuste aos dados, enquanto o RF apresentou grandes dificuldades para extrapolar o limite para qual foi treinado. Para os modelos Cubist e MARS o desempenho foi moderado e para o MLR foi mais limitado, não conseguindo boas previsões de SST. Este estudo demonstra que a utilização de sensor multiespectral acoplado a RPA apresenta elevado grau de precisão na estimativa de SST em pequenos reservatórios.

Palavras-chave: Sensoriamento remoto. Espectrometria. Água - Qualidade. Monitoramento. Imagens multiespectrais.

ABSTRACT

DIAS, Rafael Luís Silva, M.Sc., Universidade Federal de Viçosa, July, 2020. **Development of a system for monitoring the concentration of total suspended solids in reservoirs based on a multispectral sensor on board a remotely piloted aircraft and machine learning.** Adviser: Demetrius David da Silva. Co-advisers: Cibele Hummel Amaral and Elpídio Inácio Fernandes Filho.

Monitoring water quality is essential for the maintenance and conservation of water resources. The conventional method of monitoring is done by spot sampling, which is limited to representing the spatial variability of its parameters, being significantly expensive. A promising alternative is the Remote Sensing techniques with the use of orbital sensors, which have been applied to map water quality parameters, mainly Optically Active Components (OAC), such as Total Suspended Solids (TSS), Chlorophyll-a (Chl-a) and Colored Dissolved Organic Matter (CDOM). However, the monitoring of small reservoirs with orbital sensors always has limitations, such as spectral, spatial or temporal resolutions. In addition, all optical orbital sensors are subject to atmospheric disturbances, in addition to clouds, especially during rainy periods, just when the sediment transport is more intense. In this context, the objective of this work was to develop a system for monitoring the concentration of Total Suspended Solids (TSS) in reservoirs based on a multispectral sensor on board a Remote Piloted Aircraft (RPA). For this, a MicaSense RedEdge multispectral sensor was used, which has five bands in the spectra blue, green, red, rededge and infrared (NIR). Four on-site campaigns were carried out, in different seasons in 2019, varying from the drought period to the period with extreme rain conditions in 2020. Two reservoirs of water accumulation fully inserted in the Atlantic Forest biome were chosen. The data were submitted to five different machine learning algorithms: Random Forest (RF), Support Vector Machines Radial Sigma (SVM-RS), Multivariate Adaptive Regression Spline (MARS), Multiple Linear Regression (MLR) and Cubist, in order to estimate the SST concentration in the reservoirs, using the root statistical indices of the mean square error (RMSE), mean absolute error (MAE) and determination coefficient (r^2) in the comparison between the predicted and observed values. The results showed that Recursive Feature Elimination (RFE) significantly reduced the number of covariates in machine learning models. The SVM-RS model was the one that had the best fit to the data, while the RF presented great difficulties

to extrapolate the limit for which it was trained. For the Cubist and MARS models, performance was moderate and for the MLR it was more limited, not achieving good SST predictions. This study shows that the use of a multispectral sensor on board a RPA has a high degree of precision in estimating OSH in small reservoirs.

Keywords: Remote sensing. Spectrometry. Water - Quality. Monitoring. Multispectral images.

LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema representativo dos elementos fundamentais para a utilização do sensoriamento remoto óptico passivo.	20
Figura 2 - Esquema demonstrando a propagação da energia eletromagnética.....	21
Figura 3 - Representação do Espectro Eletromagnético.....	22
Figura 4 - Interações entre a radiação eletromagnética e os alvos na atmosfera.	23
Figura 5 - Espectros da água sob dominância dos COAs: total de sólidos em suspensão (SST), clorofila-a e matéria orgânica dissolvida colorida CDOM.....	28
Figura 6 - Função de resposta espectral relativa e largura total na metade do máximo do sensor MicaSense RedEdge.	32
Figura 7 - Localização dos reservatórios na bacia do córrego São Bartolomeu.....	37
Figura 8 - Fluxograma geral da metodologia criada para o desenvolvimento de sistema de monitoramento da concentração de Sólidos Suspensos Totais (SST) em reservatórios a partir de sensores multiespectrais acoplados a Aeronave Remotamente Pilotada (RPA).	38
Figura 9 - RPA DJI Matrice 100 foi usado no estudo (A) e Sensor MicaSense RedEdge e DLS (B).....	40
Figura 10 - Trajeto realizado pelo RPA no levantamento aéreo dos Reservatórios R1 e R2, representado pelas setas amarelas, e em balões verdes o momento de cada captura de imagem.....	40
Figura 11 – Distribuição dos pontos de controle no reservatório 1 (A) e no reservatório 2 (B), demarcação dos pontos de controle em alvos já existentes no local (C) e base do GNSS RTK TPS T10 instalada (D).....	41
Figura 12 - Procedimento de calibração radiométrica das imagens em campo (A) e painel de calibração radiométrica (B).	42
Figura 13 - Membranas de fibra de vidro antes (A) e após (B) o processo de filtragem do SST.	45
Figura 14 - Esquema metodológico dos processos aplicados para a estimativa de SST nos reservatórios 1 e 2.....	47
Figura 15 – Boxplot das amostras de SST nas campanhas de setembro/2019, janeiro/2020, fevereiro/2020 e março/2020 dos dois reservatórios.	50

Figura 16 - Comportamento radiométrico médio das amostras coletadas com sensor multiespectral nas campanhas de janeiro/2020 e setembro/2019.....	51
Figura 17 - Desempenho dos modelos de aprendizado de máquina no processo de seleção de variáveis para cada uma das três abordagens avaliadas.	53
Figura 18 - Comparação entre os valores observados e estimados de SST nos modelos RF, SVM-RS, MLR e MARS e Cubist nas três, aplicando diferentes abordagens.	54
Figura 19 – Importância das covariáveis na estimativa de SST utilizando o modelo RFE-SVM-RS.....	57
Figura 20 - Predição da concentração de SST pelo modelo SVM-RS para o reservatório 1 nas quatro campanhas realizadas.....	58
Figura 21 - Predição da concentração de SST pelo modelo SVM-RS para o reservatório 2 nas quatro campanhas realizadas.....	59

LISTA DE TABELAS

Tabela 1 – Comparação de sensores Multiespectrais aerotransportados com sensores abordo de satélites.....	31
Tabela 2 - Informações dos levantamentos aéreos e coleta das amostras de água realizadas nos reservatórios.....	39
Tabela 3 - Abordagens utilizadas na modelagem de SST	46
Tabela 4 – Desempenho dos modelos RF, SVM-RS, MARS, MLR e Cubist na estimativa de SST nas três diferentes abordagens	55

SUMÁRIO

1	INTRODUÇÃO.....	15
2	OBJETIVOS	18
2.1	Objetivo geral	18
2.2	Objetivos específicos	18
3	REVISÃO BIBLIOGRÁFICA	19
3.1	Sensoriamento Remoto.....	19
3.1.1	Radiação Eletromagnética (REM).....	19
3.1.2	Espectro eletromagnético	21
3.1.3	Interação da energia na atmosfera.....	22
3.1.4	Fundamentos radiométricos.....	23
3.1.4.1	Transmitância hemisférica.....	24
3.1.4.2	Reflectância hemisférica	24
3.1.4.3	Absorbância hemisférica	24
3.1.4.4	Irradiância hemisférica	24
3.1.4.5	Exitância	25
3.1.4.6	Radiância	25
3.2	Propriedades ópticas da água	25
3.2.1	Sólidos em suspensão	26
3.2.2	Fitoplânctons (pigmentos fotossintetizantes).....	26
3.2.3	Matéria orgânica dissolvida colorida (CDOM).....	27
3.3	Monitoramento da qualidade da água interior	28
3.3.1	Monitoramento da qualidade da água com sensoriamento remoto ...	29
3.3.1.1	Sensores.....	29
3.3.1.2	Sensor MicaSense	31
3.4	Aprendizado de Máquina.....	32

4	MATERIAL E MÉTODOS	36
4.1	Área de Estudo	36
4.2.	Campanhas de Campo e Levantamento Aéreo	38
4.2	Correção Geométrica	41
4.3	Calibração radiométrica	41
4.4	Ortorretificação e mosaico das imagens aéreas	42
4.5	Extração dos valores de reflectância	43
4.6	Campanhas de coleta das amostras de água	44
4.6.1	Análise das amostras	44
4.7	Levantamento batimétrico dos reservatórios	45
4.8	Modelagem de SST	46
4.8.1	Seleção das covariáveis	47
4.8.2	Seleção dos modelos de aprendizado de máquina	48
4.8.3	Avaliação dos modelos de aprendizado de máquina	48
4.9	Mapa de estimativas de SST dos Reservatórios	49
5	RESULTADOS E DISCUSSÃO	50
5.1	Amostragem de SST	50
5.2	Seleção das covariáveis	52
6	CONCLUSÕES	61
	REFERÊNCIAS	62
	APÊNDICE A – COVARIÁVEIS UTILIZADAS PARA IMPLEMENTAÇÃO DOS MODELOS	71

1 INTRODUÇÃO

As águas continentais são indispensáveis para a manutenção do ciclo de vida no planeta, pois garantem a estabilidade da biodiversidade e do fluxo hidrológico (BARBOSA; NOVO; MARTINS, 2019). Todas as grandes civilizações tiveram suas origens às margens de grandes rios, lagos ou reservatórios. Porém, o equilíbrio ecológico logo foi quebrado e, como consequência, ocasionou a degradação qualitativa e quantitativa dos recursos hídricos (PINTO, 2017). Atualmente, muitos países vivenciam cenários de escassez de água potável, considerada como um recurso estratégico do ponto de vista geopolítico no cenário internacional (TUNDISI, 2003). No Brasil a situação não é diferente, pois mesmo com a grande disponibilidade hídrica a água não é distribuída de forma igualitária, agravando conflitos entre os usuários em todas as regiões (BARBOSA; NOVO; MARTINS, 2019).

De acordo com o Relatório da Conjuntura dos Recursos Hídricos no Brasil (ANA, 2017), estima-se que a disponibilidade hídrica superficial no país seja em torno de 78.600 m³/s, o que representa 30% da vazão média de todos os rios do mundo. Entretanto, a Bacia Amazônica corresponde a 83,5% desse valor, enquanto a Bacia do Paraná (segunda maior), conta com apenas 6% da produção hídrica brasileira (BARBOSA; NOVO; MARTINS, 2019).

Em decorrência desta disparidade espacial da distribuição dos recursos hídricos, a partir do século XX ampliaram-se as exigências de um monitoramento mais rigoroso e contínuo de modo a garantir a qualidade e quantidade da água (ENGMAN; GURNEY, 1991; SLOGGETT et al., 1995). O monitoramento dos parâmetros biofísicos e químicos das águas é um meio de compreender os processos naturais e antrópicos que interferem na dinâmica natural dos sistemas hídricos (KROGER, S.; LAW, 2005).

Para analisar a qualidade da água são necessários indicadores que compreendem suas propriedades físicas, químicas e biológicas. Os parâmetros são tradicionalmente determinados a partir de coletas em campo e análises das amostras em laboratório. Embora a medição local ofereça alta precisão, é um processo oneroso e demorado, não sendo viável a obtenção de um banco de dados simultâneo de qualidade da água em escala regional (DUAN et al., 2013a, 2013b).

Além disso, os métodos convencionais de amostragem por pontos não são capazes de identificar facilmente as variações espaciais ou temporais na qualidade da água, o que é vital para uma avaliação e gerenciamento abrangentes dos corpos hídricos. Portanto, essas dificuldades de amostragens sucessivas e integradas constituem-se num obstáculo expressivo ao monitoramento e gerenciamento da qualidade da água (GHOLIZADEH; ASSEFA; REDDI, 2016).

Considerando as limitações citadas anteriormente, o uso do sensoriamento remoto e imagens de satélite têm sido amplamente difundido nos últimos anos para fins de estimativa de parâmetros de qualidade de água (ARANGO; NAIRN, 2019). Diversos autores utilizaram sensores orbitais com a capacidade de obter informações do visível (400-700nm) e infravermelho próximo - NIR (700-1000nm) do espectro eletromagnético; estes foram utilizados para a estimativa de diversos parâmetros de qualidade da água, como clorofila-a (Chl-a), turbidez (T), pH, matéria orgânica dissolvida colorida (CDOM), sólidos suspensos totais (SST) e profundidade do disco de Secchi (SSD) (ARANGO; NAIRN, 2019; BONANSEA et al., 2015a; DEKKER, 2001; GHOLIZADEH; ASSEFA; REDDI, 2016; HALABISKY et al., 2016; RITCHIE; SCHIEBE; MCHENRY, 1976; ROBERT et al., 2016; SÁENZ; PAEZ; ARANGO, 2015; SU, 2017; XU et al., 2019).

Apesar dos benefícios do uso de produtos gratuitos de sensores ópticos orbitais, eles ainda apresentam grandes gargalos na determinação de parâmetros de qualidade da água, como a temporalidade de revisita dos satélites, sua resolução espacial e a suscetibilidade excessiva à cobertura de nuvens (ASNER, 2010; DEKKER, 2001). Assim, as pesquisas tendem atualmente a ser realizadas com aeronaves remotamente pilotadas (RPA) (ARANGO; NAIRN, 2019; DOUGHTY; CAVANAUGH, 2019; SU, 2017; ZHANG et al., 2017).

Com sensores a bordo de RPA consegue-se resolução espacial centimétrica, imagens livres de nuvens e liberdade temporal para os voos, já tendo sido obtidos resultados para Chl-a e SSD em reservatórios no sudeste da China (Su, 2017), turbidez em lago no norte da Austrália (Chen, 2016) e estimativas de parâmetros ópticos e não ópticos de qualidade da água com sensor multiespectral acoplado em RPA (Arango, 2019). Estas pesquisas demonstram a relação entre as bandas espectrais e os parâmetros de qualidade da água.

Embora o sensoriamento remoto apresente excelentes resultados no monitoramento de corpos hídricos, é essencial conhecer os constituintes da água. Variáveis limnológicas como a matéria orgânica (MO), Chl-a e SST, quando relacionadas com a reflectância, possibilitam uma análise integral da amostra (KIRK, 2011).

Segundo Kirk (2011) e Jensen (2009) o SST é composto por partes orgânicas e inorgânicas, que se relacionam com a radiação eletromagnética modificando o comportamento espectral da água. Desta forma, o SST pertence à classe dos componentes opticamente ativos (COAs) que são responsáveis pelo espalhamento, retroespalhamento e absorção da radiação dentro da água. Os sólidos em suspensão têm um impacto significativo na absorção e dispersão da luz, com as partículas inorgânicas possuindo maior domínio do espalhamento e retroespalhamento, sendo responsáveis pelo aumento da reflectância na região do vermelho e do infravermelho próximo. Já a absorção pode ser dominante no caso de haver partículas orgânicas presentes na coluna de água (WILSON, 2009).

O excesso de sólidos em suspensão se torna um fator crítico para a qualidade da água, levando a um declínio nas comunidades aquáticas e a uma grave degradação ecológica dos ambientes (BILOTTA; BRAZIER, 2008). As partículas de sólidos na água afetam suas propriedades físicas e características, como densidade, viscosidade e capacidade térmica. A água turva tem capacidade térmica mais elevada, pois os sólidos em suspensão absorvem o calor da radiação solar, aumentando a temperatura da água e reduzindo os níveis de oxigênio dissolvido (DO) (PETROVI; MARKOVI, 2015).

Neste contexto, a integração entre as variáveis limnológicas e o sensoriamento remoto a partir de RPA surge como uma possibilidade para a geração e interpretação dos dados de acordo com a demanda do usuário. Esta alternativa se apresenta como ideal para estudo em pequenas áreas, como microbacias e pequenos lagos e reservatórios, onde se necessitam de dados com alta precisão e monitoramento constante (SÁENZ; PAEZ; ARANGO, 2015).

2 OBJETIVOS

2.1 Objetivo geral

Desenvolver sistema de monitoramento da concentração de Sólidos Suspensos Totais (SST) em reservatórios a partir de sensores multiespectrais acoplados a Aeronave Remotamente Pilotada (RPA).

2.2 Objetivos específicos

- i. Estabelecer procedimento para gerar medidas de reflectância a partir das imagens de sensor multiespectral;
- ii. Desenvolver modelos relacionando a reflectância das imagens aéreas às concentrações de SST utilizando aprendizado de máquina;
- iii. Avaliar a influência das condições do ambiente na estimativa da concentração de SST; e
- iv. Gerar mapa com a espacialização das estimativas de SST.

3 REVISÃO BIBLIOGRÁFICA

3.1 Sensoriamento Remoto

O conceito de sensoriamento remoto é muito amplo e existem várias definições para cada área da ciência. Na literatura, foi utilizado pela primeira vez em 1960 com referência à aquisição de informações sem contato direto com o objeto alvo, sendo esta a primeira definição de sensoriamento remoto. A partir de então vem sendo utilizado para diferentes campos da tecnologia e conhecimentos variados, como a física, botânica, engenharia, eletrônica e cartografia. Novo (2010) definiu sensoriamento remoto como sendo a utilização conjunta de sensores, equipamentos para processamento de dados, equipamentos de transmissão de dados colocados a bordo de aeronaves, espaçonaves ou outras plataformas, com o objetivo de estudar eventos, fenômenos e processos que ocorrem na superfície do planeta Terra, a partir do registro e da análise das interações entre a radiação eletromagnética e as substâncias que o compõem em suas diversas manifestações (NOVO, 2008).

Nos últimos anos, o sensoriamento remoto vem se expandindo exponencialmente nos variados campos da ciência (BARBOSA; NOVO; MARTINS, 2019). No monitoramento aquático não é diferente, com diversos autores se dedicando à investigação das potencialidades de utilização do sensoriamento remoto para a previsão e estimativa de parâmetros ambientais. Sáenz, Paez e Arango (2015) utilizaram drones para o monitoramento de SST em bacias hidrográficas, Kim et al. (2017) determinou a eficiência do sensor MODIS na determinação de Chl-a e SST. Já Su (2017) construiu algoritmos empíricos para a estimativa de parâmetros de qualidade de água com RPA.

3.1.1 Radiação Eletromagnética (REM)

A fonte de energia para todos os processos físico-químicos e biológicos que ocorrem na superfície terrestre é o sol. A radiação solar é também a fonte geratriz de todos os processos atmosféricos (TUBELIS, A.; NASCIMENTO, 2000). Ao interagir com os materiais que compõe a superfície terrestre produz radiação de retorno que irá se propagar pela atmosfera e atingir o sensor. O que chega até o sensor é uma

certa intensidade de energia eletromagnética (radiação), que será posteriormente transformada em um sinal passível de interpretação, a radiância (JENSEN, 2009; SANTOS, 2013). Na Figura 1 são apresentados os elementos fundamentais para a aplicação do sensoriamento remoto.

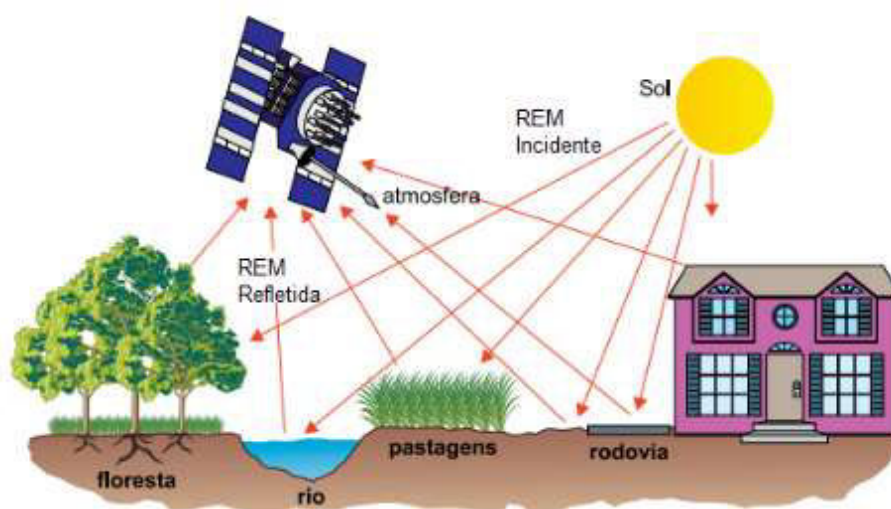


Figura 1 - Esquema representativo dos elementos fundamentais para a utilização do sensoriamento remoto óptico passivo.

Fonte: adaptado de IBGE, 2016.

Como demonstrado na Figura 1, a REM é o elemento de ligação entre todos os demais elementos fundamentais: fonte (Sol), alvo (superfície terrestre) e sensor (instrumento capaz de coletar os dados). A fonte de REM fundamental para a aplicação das técnicas de sensoriamento remoto no estudo dos recursos naturais é o Sol, o sensor é o instrumento capaz de coletar e registrar a REM refletida ou emitida pelo alvo (que representa o elemento do qual se pretende extrair informação) (NOVO, 2010).

A quantidade da REM refletida pelos objetos terrestres resulta das interações entre ela e os componentes físicos, químicos e biológicos destes objetos. Essas interações são passíveis de identificação em imagens de sensores remotos (MORAES, 2002a)

Duas características da REM são importantes para entender os princípios do Sensoriamento Remoto, sendo elas o comprimento e a frequência de onda. O comprimento de onda é determinado pela distância entre duas cristas sucessivas de ondas, sendo medido em metros (m) ou em seus submúltiplos. A frequência de onda

refere-se ao número de ondas que passam por um ponto fixo por unidade de tempo, sendo medida em *Hertz* (Hz), o que equivale a um ciclo por segundo (SILVA, 2006). Esses elementos são apresentados na Figura 2.

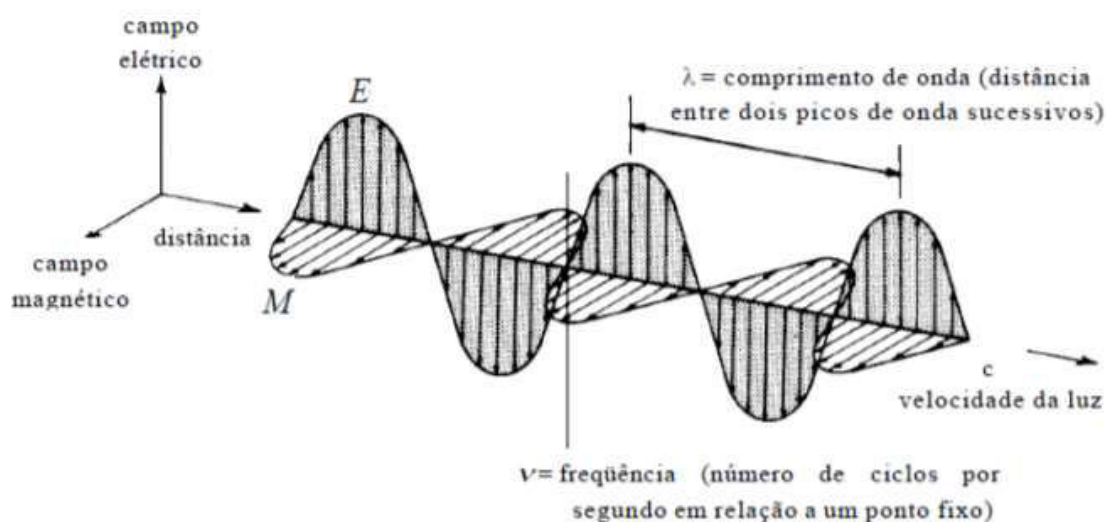


Figura 2 - Esquema demonstrando a propagação da energia eletromagnética.
Fonte: Adaptado de Lillesand e Kiefer (1994).

3.1.2 Espectro eletromagnético

O conjunto de todas as radiações conhecidas, desde os raios cósmicos e gama (com alta frequência e pequenos comprimentos de ondas) até as ondas de rádio (com baixa frequência e grandes comprimentos de onda), é denominado espectro eletromagnético (Figura 3). O espectro eletromagnético é a ordenação destas radiações em função do comprimento de onda e da frequência (MOREIRA, 2011).

A medida que se avança para a direita do espectro eletromagnético as ondas apresentam maiores comprimentos de onda e menores frequências (Figura 3). A faixa espectral mais utilizada em sensoriamento remoto estende-se de $0,3 \mu\text{m}$ a $15 \mu\text{m}$, embora a faixa de microondas também seja utilizada.

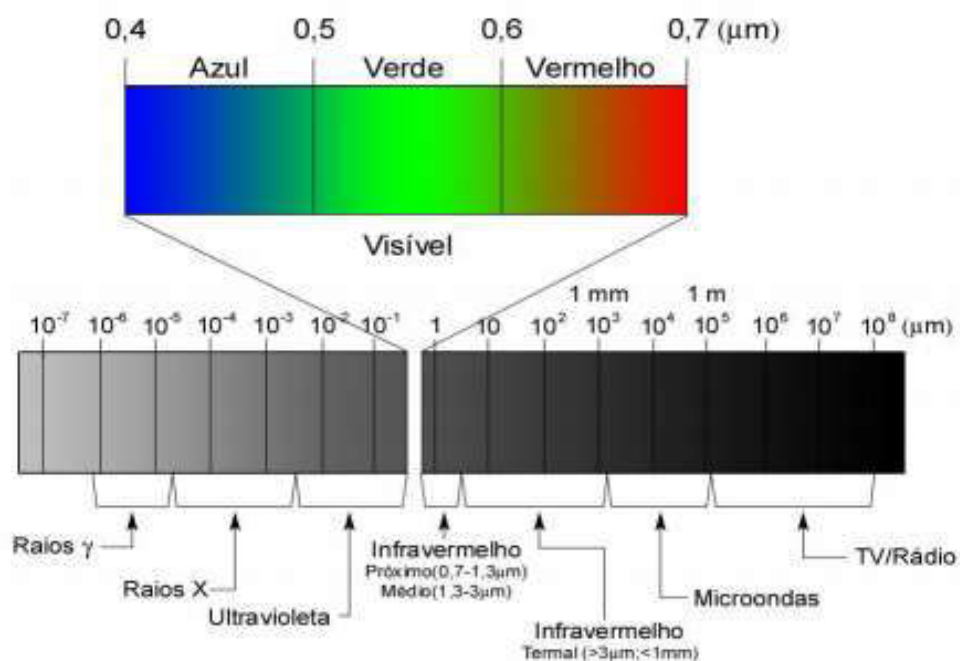


Figura 3 - Representação do Espectro Eletromagnético.
Fonte: Moreira (2011).

3.1.3 Interação da energia na atmosfera

A radiação eletromagnética é propagada através do vácuo e da atmosfera até atingir a Terra. A atmosfera é formada por uma mistura de gases, vapor de água e impurezas. Quando a energia eletromagnética incide sobre um alvo da superfície terrestre, três interações fundamentais ocorrem entre a energia e o alvo: pode ser parcialmente refletida, absorvida e transmitida, conforme apresentado na Figura 4, sendo que as propriedades intrínsecas de cada objeto é que definem a percentagem de cada fração em que o fluxo é subdividido (LILLESAND, T. M.; KIEFER, 1994).

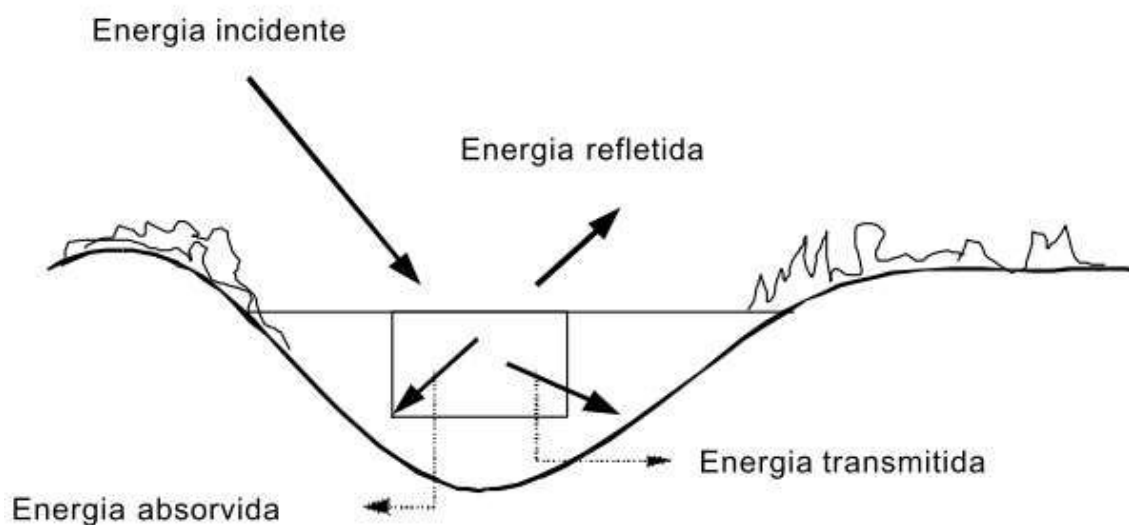


Figura 4 - Interações entre a radiação eletromagnética e os alvos na atmosfera.
 Fonte: Adaptada de Lillesand e Kiefer (1994).

A Refração refere-se ao fenômeno pelo qual a energia é transmitida por um determinado material em decorrência da diferença de densidade entre os meios pelos quais a radiação eletromagnética se propaga. O Espalhamento é definido pela reflexão da energia, assim é o fenômeno pelo qual ocorre a difusão de forma elevatória da radiação por partículas da atmosfera. Já a Absorção é o processo pelo qual a energia radiante é absorvida e convertida em outras formas de energia (SEYHAN; DEKKER, 1986).

3.1.4 Fundamentos radiométricos

Para o entendimento das relações da REM com a superfície terrestre, é necessário estimar algumas medidas que quantifiquem a energia produzida pela fonte, a quantidade que foi atenuada pelo meio que se interpõe entre a fonte e o alvo e a quantidade de energia que o alvo absorveu (NOVO, 2010). Para compreender melhor os fundamentos radiométricos é necessário entender as métricas utilizadas para quantificar a energia que é transmitida, refletida ou absorvida pelas superfícies, conforme definido por Nicodemus (1970):

3.1.4.1 Transmitância hemisférica

A transmitância hemisférica (τ_λ) é definida como a razão adimensional entre o fluxo radiante transmitido por uma superfície e o fluxo radiante incidente sobre ela, representada pela equação 1.

$$\tau_\lambda = \frac{\phi_{\text{transmitido } \lambda}}{\phi_{i\lambda}} \quad (1)$$

3.1.4.2 Reflectância hemisférica

A reflectância hemisférica (ρ_λ) é definida como a razão adimensional entre o fluxo radiante refletido por uma superfície e o fluxo radiante incidente sobre ela, multiplicada por 100 se obtém a reflectância da superfície em porcentagem ($\rho_{\lambda\%}$), representada pela equação 2.

$$\rho_\lambda = \frac{\phi_{\text{refletido } \lambda}}{\phi_{i\lambda}} \times 100 \quad (2)$$

3.1.4.3 Absorbância hemisférica

A absorbância hemisférica (α_λ) é definida como a razão adimensional entre o fluxo radiante absorvido por uma superfície e o fluxo radiante incidente sobre ela, representada pela equação 3.

$$\alpha_\lambda = \frac{\phi_{\text{absorvido } \lambda}}{\phi_{i\lambda}} \quad (3)$$

3.1.4.4 Irradiância hemisférica

A irradiância hemisférica (E_λ) é a quantidade de fluxo radiante incidente sobre uma superfície por unidade de área da superfície, medida em *Watts* por metro quadrado (W m^{-2}), como apresentado na equação 4.

$$E_\lambda = \frac{\phi_\lambda}{A} \quad (4)$$

3.1.4.5 Exitância

A exitância (M_λ) é a quantidade de fluxo radiante que emerge de uma superfície por unidade de área da superfície, medida em *Watts* por metro quadrado (W m^{-2}), conforme equação 5:

$$M_\lambda = \frac{\phi_\lambda}{A} \quad (5)$$

3.1.4.6 Radiância

A radiância (L_λ) é a intensidade radiante por unidade de área projetada numa direção específica. É medida em *Watts* por metro quadrado por estereoradiano. O fluxo radiante em certos comprimentos de onda (ϕ_λ), deixando a área-fonte projetada (A) numa certa direção ($\cos \theta$) e num ângulo sólido (Ω), é demonstrado na equação 6.

$$L_\lambda = \frac{\frac{\phi_\lambda}{\Omega}}{A \cos \theta} \quad (6)$$

3.2 Propriedades ópticas da água

As diversas misturas de partículas e materiais dissolvidos ou em suspensão na água, que afetam processos de absorção e de espalhamento da luz, são denominadas componentes ou constituintes opticamente ativos (COAs). Alguns destes materiais interagem com a radiação incidente, modificando as respostas espectrais e, conseqüentemente, a cor da água. Os COAs são: a água pura, a matéria orgânica dissolvida colorida (CDOM) e o particulado total (fitoplâncton e sedimentos minerais e orgânicos). Cada um desses componentes têm propriedades ópticas específicas de absorção e de espalhamento da radiação solar, ou seja, têm propriedades ópticas que lhes são inerentes (KIRK, 2011).

Como os COAs podem coexistir em diferentes proporções e interagir simultaneamente com a radiação solar, os efeitos dos processos de absorção/espalhamento de um COA interferem nos efeitos dos processos de outro COA. Dessa forma, os resultados das interações simultâneas dos COAs com a luz num corpo d'água determinam o comportamento espectral das massas de água e

suas variações de concentrações regulam a intensidade da radiação retro espalhada pela coluna d'água (BARBOSA; NOVO; MARTINS, 2019).

3.2.1 Sólidos em suspensão

A concentração dos sedimentos em suspensão é o parâmetro de qualidade de água que afeta com maior intensidade o comportamento óptico da mesma. Os sedimentos em suspensão são compostos por partículas orgânicas (detritos e microrganismos vivos) e inorgânicas (minerais), presentes na coluna d'água. As partículas inorgânicas variam desde a argila fina (3 - 4 μm em diâmetro) até a areia grosseira (131 - 250 μm) e as características dessas partículas afetam a reflectância da água de forma direta. A soma da matéria orgânica morta e matéria inorgânica em suspensão é referenciada como *tripton* ou fração não algal (BARBOSA, 2005; BARBOSA; NOVO; MARTINS, 2019).

Os sedimentos em suspensão são originários de diversas fontes, incluindo erosão de áreas agrícolas, intemperismo de terrenos montanhosos, erosões costeiras, entre outras. O estudo e monitoramento do tipo e quantidade desses sedimentos são importantes. Um exemplo disso é a erosão do solo na bacia hidrográfica, que lança cargas de sedimentos nos corpos d'água, resultando no assoreamento mais rápido dos principais rios e reservatórios, além de afetar a qualidade da água e suas aplicações para o consumo humano (JENSEN, 2009).

Com o aumento da concentração de sedimentos em suspensão, da turbidez da água e do aumento geral da reflectância, ocorre modificação da forma do espectro de reflectância. Porém, a radiação que sofre espalhamento devido à concentração de sedimentos não tem um comportamento constante e varia em função dos tamanhos das partículas e do material de origem, o que torna difícil a formulação de um algoritmo geral para os sedimentos em suspensão (NOBREGA, 2002).

3.2.2 Fitoplânctons (pigmentos fotossintetizantes)

Os principais pigmentos que contribuem para a absorção da radiação solar são os fotossintetizantes, como clorofila e carotenoides. A concentração da clorofila-a (chl-a) é usada como um indicador da presença de fitoplâncton e do estado trófico de um sistema aquático (BARBOSA; NOVO; MARTINS, 2019).

A variação dos níveis de clorofila indica a floração de diferentes algas e cianobactérias; o agente causador da eutrofização em água doce está diretamente relacionado à concentração de clorofila, uma vez que esse pigmento é essencial para a fotossíntese (LIM; CHOI, 2015). Essas concentrações podem variar dependendo do metabolismo, luz, temperatura, disponibilidade de nutrientes e outros fatores. Portanto, o registro da cor de uma área do corpo d'água possibilita estimar a quantidade e tipo de fitoplâncton presente (JENSEN, 2009).

A interpretação das feições de absorção desse pigmento dependerá de bandas espectrais na região do azul (450 nm) e vermelho (670 nm), enquanto o pico de reflectância ocorre na região do verde (550 nm) e início do NIR (710 nm). Em muitos estudos, a aplicação da razão de bandas (ex.: 710/670 nm) se torna útil pelo forte contraste entre espalhamento e absorção nesses comprimentos de ondas (BARBOSA; NOVO; MARTINS, 2019).

3.2.3 Matéria orgânica dissolvida colorida (CDOM)

A CDOM é composta por substâncias orgânicas (ácidos húmicos e fúlvicos) dissolvidas que alteram a cor da água. Essas substâncias geram tons de amarelo até marrom escuro na água devido à forte absorção na região do visível, principalmente na faixa do azul. Conforme a magnitude e composição, a CDOM afeta os processos fotoquímicos e regem o campo de luz subaquático. Além disso, a CDOM é fonte de nutrientes para as plantas subaquáticas, em consequência dos compostos gerados da biodegradação da matéria orgânica (BARBOSA; NOVO; MARTINS, 2019).

A absorção da CDOM aumenta exponencialmente em comprimentos de ondas menores, como 380 e 500 nm e, normalmente, esse constituinte é caracterizado pelo coeficiente de absorção espectral em 410 – 440 nm. Em razão da sobreposição das feições de absorção da clorofila-a e da CDOM em comprimentos de ondas entre 380 e 440 nm, a estimativa desse constituinte passa a representar um desafio para os estudos. Muitos dos algoritmos baseiam-se em regressões empíricas via reflectância em comprimentos de onda acima de 500 nm (ZHU et al., 2014). Na Figura 5 pode-se observar a influência dos COAs no espectro da água.

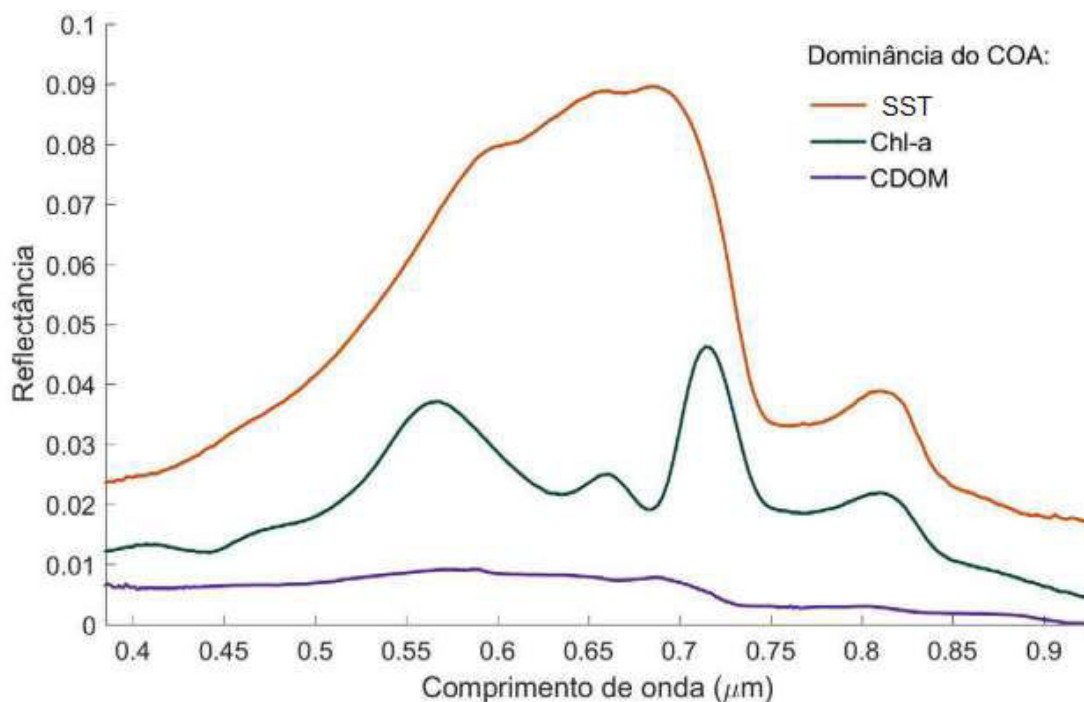


Figura 5 - Espectros da água sob dominância dos COAs: total de sólidos em suspensão (SST), clorofila-a e matéria orgânica dissolvida colorida CDOM.

Fonte: Barbosa, Novo e Martins (2019).

3.3 Monitoramento da qualidade da água interior

A qualidade da água é geralmente quantificada através da avaliação de indicadores, incluindo indicadores químicos, tais como: oxigênio dissolvido (OD), demanda química de oxigênio (DQO), demanda bioquímica de oxigênio (DBO), metais pesados tóxicos, fósforo e nitratos, condutividade elétrica, alcalinidade total; indicadores físicos, tais como: temperatura, pH, sólidos suspensos totais (SST), turbidez, sólidos totais dissolvidos (STD; e indicadores biológicos, tais como: a *Escherichia coli* (E. coli) de bactérias fecais e clorofila-a (BERKMAN; CANOVA, 2007; GIBS; WILDE; HECKATHORN, 2012).

Os atuais programas de monitoramento ambiental têm como origem coletas pontuais em campo. As amostras coletadas são enviadas para o laboratório onde são analisadas, sendo o resultado da análise utilizado para inferir a qualidade de água de toda uma diversidade (BERKMAN; CANOVA, 2007). Embora a medição *in situ* ofereça alta precisão, é um processo oneroso e demorado, não sendo viável para fins de fornecimento de um banco de dados simultâneo de qualidade da água em escala regional (DUAN et al., 2013b).

3.3.1 Monitoramento da qualidade da água com sensoriamento remoto

Com a utilização de sensores remotos é possível avaliar grandes e pequenas áreas, fornecendo valiosos dados de padrões complexos dos sistemas aquático, o que é difícil de conseguir com amostras em campo nos monitoramentos tradicionais da qualidade da água.

Dentre as vantagens da utilização de sensores remotos comparativamente ao monitoramento tradicional tem-se: o menor custo operacional, maior agilidade nos resultados e a possibilidade de se analisar a distribuição espacial dos dados em condições diferentes (GHOLIZADEH; ASSEFA; REDDI, 2016). Porém, Heinonen, Ziglio e Van der Beken (2008) salientam que uso integrado de sensoriamento remoto, medições no local e modelagem computacional da qualidade da água podem levar a um conhecimento maior da qualidade da água dos sistemas hídricos.

3.3.1.1 Sensores

Diversos sensores são utilizados para medir a quantidade de radiação em vários comprimentos de onda refletidos pela superfície da água. Essas reflexões podem ser usadas direta ou indiretamente para detectar diferentes indicadores de qualidade da água (SEYHAN; DEKKER, 1986).

Os sensores de observação são divididos em duas categorias principais, com base nas plataformas em que estão localizados. Sensores aéreos são aqueles que são montados em uma plataforma dentro da atmosfera da Terra (balões, helicóptero, RPAs). Sensores de espaçonave são transportados por uma espaçonave ou satélite para locais fora da atmosfera da Terra. É necessário entender as propriedades desses sensores para escolher o mais apropriado para os objetivos do estudo (GHOLIZADEH; ASSEFA; REDDI, 2016).

A qualidade de um sensor geralmente é especificada pela sua capacidade de obter medidas detalhadas da energia eletromagnética. As características dos sensores estão relacionadas com as resoluções espacial, espectral e radiométrica.

A resolução espacial representa a capacidade do sensor distinguir objetos, indicando o tamanho do menor elemento da superfície individualizado pelo sensor. Essa resolução depende principalmente do detector, da altura e do posicionamento do sensor em relação ao objeto. Para um dado nível de posicionamento do sensor, quanto menor for sua resolução geométrica, maior será o grau de distinção entre

objetos próximos. Por exemplo, o sistema sensor do *Thematic Mapper* (TM) do Landsat 5 possui uma resolução espacial de 30 metros (MORAES, 2002b).

A resolução espectral refere-se à largura espectral em que opera o sensor. Portanto, ela define o intervalo espectral no qual são realizadas medidas, e, conseqüentemente, a composição espectral do fluxo de energia que atinge o detector. Quanto maior for o número de medidas num determinado intervalo de comprimento de onda, melhor será a resolução espectral da coleta. Por exemplo, o Landsat 5 possui os sensores TM e *Multispectral Scanning System* (MSS). O sensor TM apresenta algumas bandas espectrais mais estreitas do que o sensor MSS e, portanto, nestas bandas, o TM apresenta melhor resolução espectral do que o MSS (FLORENZANO, 2002).

A resolução radiométrica define a eficiência do sistema em detectar pequenos sinais, ou seja, refere-se à maior ou menor capacidade do sistema sensor em identificar e registrar diferenças na energia refletida e/ou emitida pelos elementos que compõe a cena (rochas, solos, vegetações, águas, etc.). Por exemplo, o sistema sensor TM do Landsat 5 distingue até 256 tons distintos de sinais, representando-os em 256 níveis de cinza (NOVO, 2010)

Uma outra qualidade importante é a resolução temporal do sensor, que está relacionada com a repetitividade com que o sistema sensor pode adquirir informações referentes ao objeto. Por exemplo, os sensores do Landsat 5 possuem uma repetitividade de 16 dias (BARBOSA; NOVO; MARTINS, 2019)

Na Tabela 1 são apresentados dados detalhados sobre o comparativo entre o sensor multiespectral *MicaSense* a bordo de um RPA e sensores remotos acoplados em satélites.

Tabela 1 – Comparação de sensores Multiespectrais aerotransportados com sensores abordo de satélites

	RPA com <i>MicaSense</i>	Satélites de sensoriamento remoto
Plataforma	RPA	Terra e Água/Landsat 1 a 8/ Sentinel 2
Sensores	<i>MicaSense</i> multiespectral	MODIS, Thematic Mapper, Operatin Land Imager, MSI
Altitude/ Orbita	Até 450 m	~ 700 km
Resolução espacial	1 cm a 1m	10 m a 1 km
Custo	Baixo custo para pequenas áreas	baixo custo ou grátis
Largura da imagem	Até 1 km	16 km a 2.200 km
Resolução temporal	Ilimitada	Baixa
Escala de tempo	Ilimitada	Dias ou meses
Resolução radiométrica	12 bits	8 a 16 bits
Banda das Imagens	R/G/B/ <i>RedEdge</i> /NIR	Multiespectral (3 a 13 bandas) Hiperespectral (> 100 bandas)
Base de dados	N/A	A partir de 1984
Limitações	Instabilidade em dias de vento/Falta de algoritmos robustos para qualidade da água	Dependência climática/baixa resolução temporal

Fonte: Adaptado de Chen (2016).

3.3.1.2 Sensor *MicaSense*

O sensor multiespectral *MicaSense RedEdge* é sensível a energia refletida em 5 bandas do espectro eletromagnético, que possuem centro de banda (Full Width at Half Maximum – FWHM) variável de 475 (B1 - Azul), 560nm (B2 - Verde), 668nm (B3 - Vermelho), 717nm (B5 – *rededge*), e 840nm (B4 - infravermelho próximo-NIR) (Figura 6), tem distância focal de 5,5 mm e um campo de visada de 47,2°. O semicondutor de óxido metálico complementar (CMOS) é de 1280 × 960 pixels por quadro. A resolução radiométrica é de 12 bits, embora a saída esteja no formato TIFF-16 bits (MICASENSE, 2020).

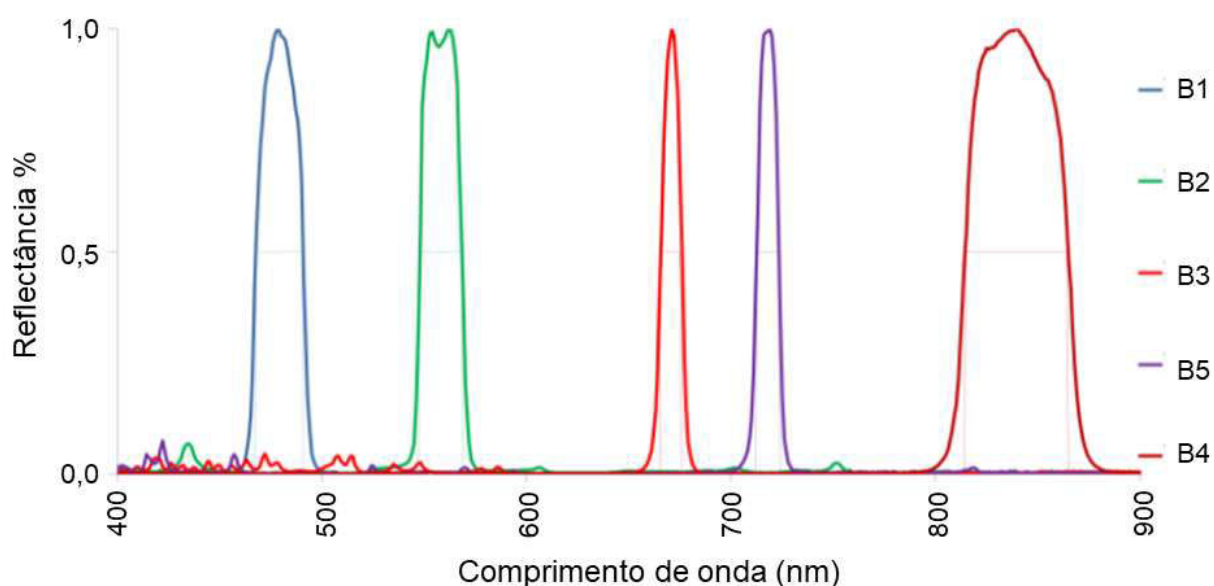


Figura 6 - Função de resposta espectral relativa e largura total na metade do máximo do sensor MicaSense RedEdge.

Fonte: Adaptado de MicaSense (2020).

3.4 Aprendizado de Máquina

Segundo Murphy (2013), o aprendizado de máquina é definido como um conjunto de técnicas que permite o aprendizado ou a identificação de padrões presentes nos dados. Com essas informações, é possível tentar prever eventos futuros ou auxiliar na tomada de decisão em cenário incerto. Assim, os métodos propostos por esta área de conhecimento são ferramentas muito úteis para resolver problemas como classificações, regressões e previsões

Os problemas de aprendizado de máquina podem ser divididos em 3 categorias, com base no tipo de aprendizado ou adaptação, segundo Kuhn e Johnson (2013):

1. Aprendizado supervisionado – também conhecido como aprendizado preditivo. Nesse caso, de dados de treinamento (T) contendo n amostras de pares de entrada (X_i) e saída (y_i) do formato $T = \{(X_i, y_i)\} \quad n_i = 1$, se procura extrair os padrões que mapeiam a saída;
2. Aprendizagem não supervisionada – neste caso, não há definição anterior de classes que identifiquem a categoria ou o tipo de padrão nos dados. Tem como objetivo encontrar "padrões interessantes ou naturais" nos dados;
3. Aprendizado por Reforço – ou aprendizado com um crítico, é baseado em sinais de recompensa ou punição por fornecer *feedback* quando a categoria

provisória é certa ou errada, respectivamente. Assim, nenhum sinal de categoria desejado é dado.

Embora façam parte de uma mesma tecnologia, os algoritmos de aprendizado de máquina são construídos por diferentes sistemas e podem apresentar funções diversificadas. Entre os sistemas podem ser mencionados os algoritmos de árvore de decisão, regressões lineares e não lineares, *support vector machines*, regressões lógicas, *naive bayes* e os algoritmos de *clustering* (KUHN; JOHNSON, 2013). A diferença entre os sistemas está na forma como é realizado o gerenciamento dos dados em seu conjunto de treinamento (LABOISSIERE; FERNANDES; LAGE, 2015).

O algoritmo *Random Forest* (RF) se enquadra como um algoritmo de árvore, foi desenvolvido para realizar classificação e regressão. Baseia-se no conceito de árvores de regressão e classificação, isto é, uma série de regras de decisão aninhadas para os preditores que determinam a resposta (BREIMAN, 2001). Tem como características a capacidade de reduzir o ruído experimental e aprimorar a precisão da previsão, agregando várias previsões diferentes. Ele cria árvores repetidamente a partir de amostras aleatórias dos dados de treinamento. Cada árvore é um modelo separado do conjunto. As estimativas de todas as árvores são finalmente calculadas para produzir a estimativa final. Para superar a correlação entre árvores, apenas um subconjunto de preditores (m try) é selecionado aleatoriamente em cada divisão. O melhor preditor do subconjunto aleatório é usado na respectiva divisão para particionar os dados (MEYER et al., 2018).

O *Support Vector Machine* - kernel da função de base radial (SVM-RS) constitui-se em regressões não lineares normalmente obtidas pela aplicação do chamado truque do kernel. Para isso, os dados de treinamento são primeiramente transformados em um espaço de maior dimensão, aplicando uma função não linear do kernel (SHAWE-TAYLOR; CRISTIANINI, 2004). Em seguida, um modelo linear é adaptado a esse novo espaço de recurso (OLSON; DELEN, 2008) e, finalmente, a regressão linear no novo espaço de recurso é equivalente a uma regressão não linear no espaço preditor original.

O algoritmo estatístico MARS foi desenvolvido como um método estatístico para ajustar o relacionamento entre um conjunto de variáveis de entrada e variáveis dependentes (FRIEDMAN, 1991). É um método de regressão não linear e não

paramétrico, baseado em uma estratégia de divisão e conquista, na qual os conjuntos de dados de treinamento são particionados em segmentos lineares separados por partes (*splines*) de gradientes diferentes (inclinação). Nenhuma suposição específica sobre o relacionamento funcional subjacente entre as variáveis de entrada e saída é necessária. Os pontos finais dos segmentos são chamados de nós. Um nó marca o final de uma região de dados e o início de outra (HOSSEINZADEH et al., 2020). O MARS seleciona automaticamente quais variáveis usar (algumas variáveis são importantes, outras não) e os locais dos nós para cada variável escolhida. As curvas por partes resultantes (conhecidas como funções básicas) oferecem maior flexibilidade ao modelo, permitindo dobras (ZHANG, 2020).

A *Multiple Linear Regression* (MLR) é a técnica de previsão mais amplamente usada para definir a relação entre as variáveis dependentes (resposta) e independentes (explicativas) (BEYCA et al., 2019). A relação entre as variáveis dependentes e independentes é formulada como um modelo de regressão múltipla na Equação 7.

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (7)$$

Onde y_i é a variável dependente, $X_{1i}, X_{2i}, \dots, X_{pi}$ são as variáveis independentes, $\beta_0, \beta_1, \dots, \beta_p$ são os coeficientes de regressão e ε é o resíduo. Os coeficientes de regressão são obtidos pelo método do erro do quadrado mínimo. O termo de erro, ε , é considerado uma variável aleatória independente e normalmente distribuída com média de zero e variância (constante) de σ^2 .

O Algoritmo *Cubist* é um modelo baseado em regras criado como uma extensão do modelo de árvore M5 (ADAMS e STERLING, 1992). Sua estrutura consiste em um componente condicional - ou função por partes - atuando como uma árvore de decisão, juntamente com vários modelos de regressão linear (KUHN, 2020). É semelhante a um modelo típico de árvore de regressão em termos de ser um algoritmo de particionamento de dados que permite explorar relações não lineares nos dados observados. O modelo particiona recursivamente os dados em subconjuntos, que são mais homogêneos internamente em relação à variável de destino e covariáveis do que o conjunto de dados como um todo. Um modelo de regressão é ajustado para cada regra com base no subconjunto de dados definido

pelas regras do modelo. Uma série de regras, organizadas hierarquicamente, define o particionamento recursivo da variável de previsão, a fim de minimizar o desvio padrão em todas as possíveis divisões (NOI; DEGENER; KAPPAS, 2017).

Observando as diferentes características de cada modelo, diversos autores realizaram estudos buscando avaliar e prever parâmetros ambientais. Noi, Degener e Kappas (2017) comprovaram a eficácia do algoritmo RF nas previsões de dados ambientais com conjunto de dados aleatórios; Hengl et al. (2018) obtiveram excelentes resultados utilizando o algoritmo RF para modelagem preditiva de variáveis espaço e espaço-temporais. Sharaf El Din, Zhang e Suliman (2017) encontraram ótimos resultados realizando a previsão de parâmetros de qualidade da água com SVM; Najafzadeh e Zainolabedini (2019) avaliaram o desempenho de estações de qualidade de água com SVM e diversos outros modelos. Já Filgueiras et al. (2019) observaram que o algoritmo *Cubist* apresenta ótimos resultados para monitoramento de colheita, indicando ainda o algoritmo MLR com bons resultados para o mesmo conjunto de dados.

4 MATERIAL E MÉTODOS

4.1 Área de Estudo

Dois pequenos reservatórios de acumulação de água foram escolhidos para a amostragem, ambos localizados na bacia do córrego São Bartolomeu, pertencente à bacia do rio Doce, que se situa na região Sudeste do Brasil, a mais populosa do país e de grande importância econômica, detendo 55,4% do PIB nacional (IBGE, 2010).

Denominados como R1 e R2, possuem 3,25 e 3,56 hectares de espelho água, respectivamente (Figura 7), estando localizados entre os paralelos 20°45'95" e 20°46'62" S e os meridianos 42°52'11,65" e 42°52'15,53" W. O reservatório R2 é responsável por parte do abastecimento de água da cidade de Viçosa-MG, com uma população estimada de 78.846 pessoas (IBGE, 2019).

A bacia do córrego São Bartolomeu apresenta declividade média de 13%, com altimetria variando de 630 a 850 metros e relevo predominantemente forte ondulado e montanhoso (EMBRAPA, 1979). Em relação às classes de solo, predominam os Latossolos Vermelho Amarelo e Argissolos Vermelho Amarelo (EMBRAPA, 2006). Segundo a classificação climática de Köppen, o clima é Cwa, clima subtropical úmido com inverno seco e verão quente (ALVARES et al., 2013).

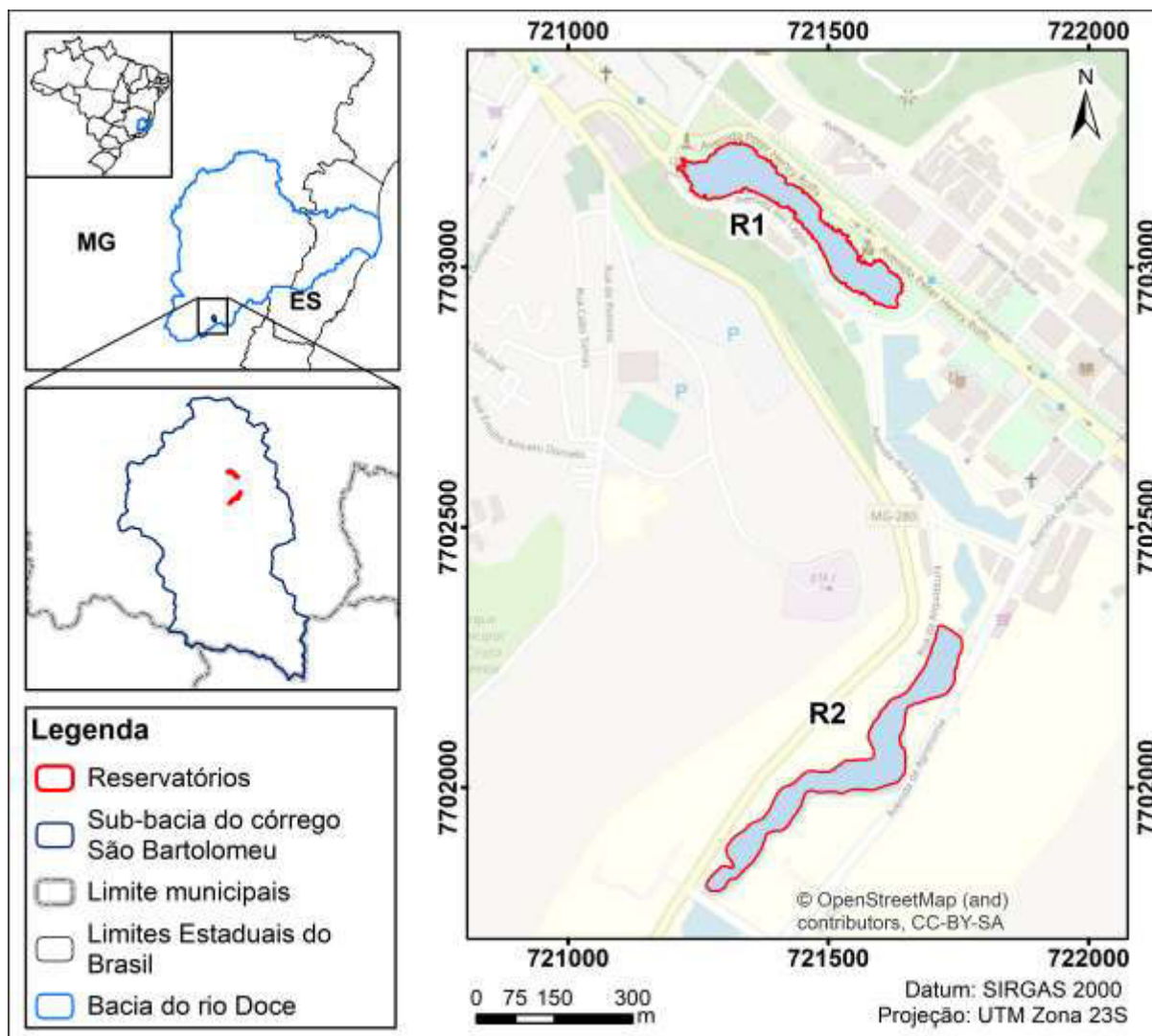


Figura 7 - Localização dos reservatórios na bacia do córrego São Bartolomeu.
 Fonte: autoria própria.

Na Figura 8 apresenta-se o fluxograma da metodologia criada para o desenvolvimento de sistema de monitoramento da concentração de Sólidos Suspensos Totais (SST) em reservatórios de acumulação de água a partir de sensores multiespectrais acoplados a Aeronave Remotamente Pilotada (RPA). A identificação do comportamento espectral do SST com sensores espectrais foi realizada nos dois reservatórios no período de setembro de 2019 a março de 2020. As imagens foram coletadas através de um sensor multiespectral MicaSense RedEdge acoplado a um RPA DJI *Matrice* 100, sendo calibradas com painel específico do sensor multiespectral para representar a verdadeira propriedade de reflectância superficial. As amostras de água foram coletadas *in loco* e sua localização demarcada com GNSS (*Global Navigation Satellite System*) RTK (*Real*

Time Kinematic); posteriormente, foram levadas para laboratório para realização das análises limnológicas. Os dados de reflectância e SST foram analisados através de algoritmos de aprendizado de máquina, sendo escolhido o modelo com melhores resultados dos índices estatísticos (RMSE, MAE e r^2) para gerar os mapas com a espacialização de SST em toda a área do reservatório.

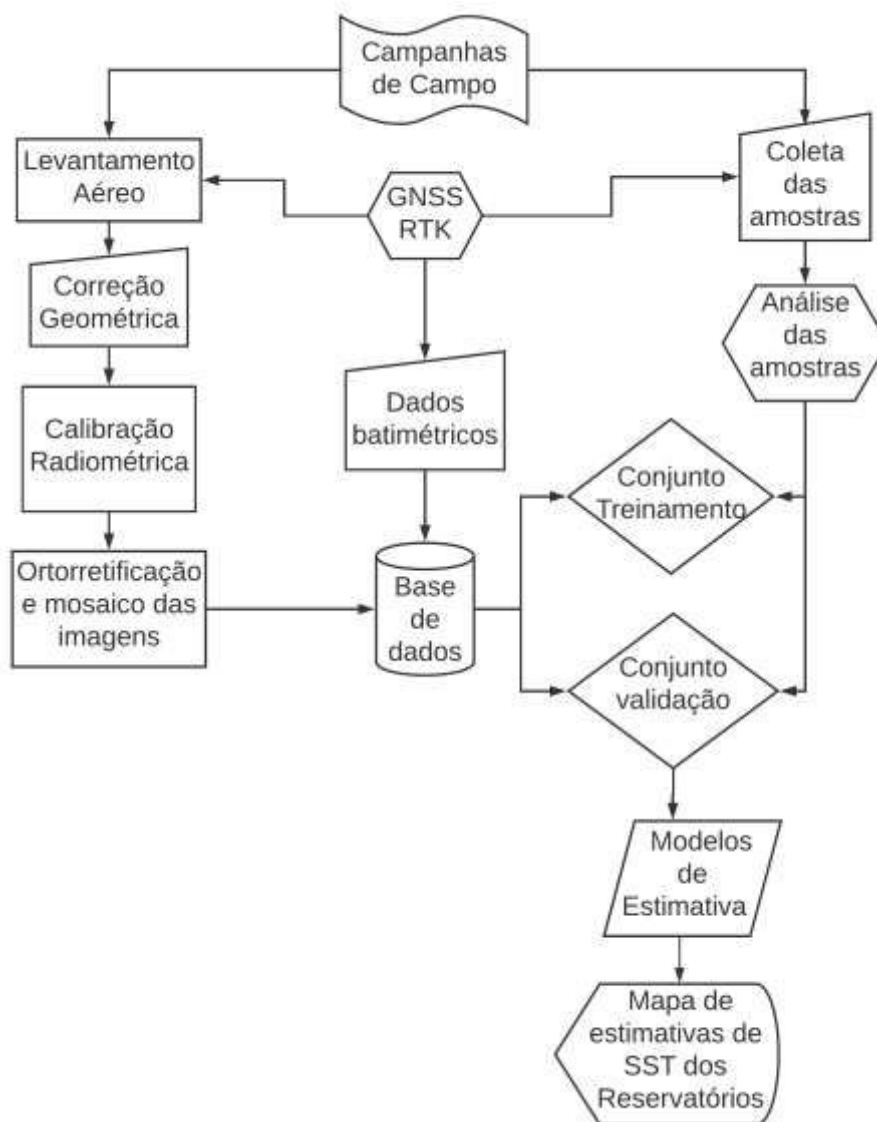


Figura 8 - Fluxograma geral da metodologia criada para o desenvolvimento de sistema de monitoramento da concentração de Sólidos Suspensos Totais (SST) em reservatórios a partir de sensores multiespectrais acoplados a Aeronave Remotamente Pilotada (RPA).

Fonte: autoria própria.

4.2. Campanhas de Campo e Levantamento Aéreo

As campanhas de campo foram divididas de modo a se avaliar desde o período de estiagem (setembro) até o período com chuvas intensas (janeiro), onde concentração de SST é maior devido ao carreamento dos sedimentos pelo

escoamento superficial. Assim, foram realizadas quatro campanhas de campo em cada reservatório (**Erro! Fonte de referência não encontrada.**). Na aquisição dos dados aéreos foi utilizado um Sistema de Aeronave Remotamente Pilotada (RPAs), integrando por um RPA DJI *Matrice* 100 e um sensor multiespectral *MicaSense RedEdge* acoplado na parte inferior (Figura 9).

Tabela 2 - Informações dos levantamentos aéreos e coleta das amostras de água realizadas nos reservatórios.

Campanhas	Local	Data	Hora de início das coletas de amostras de água	Hora de início dos voos
1	R1	10/09/2019	9:40	12:25
1	R2	11/09/2019	9:56	11:34
2	R1	24/01/2020	9:32	11:53
2	R2	25/01/2020	10:03	11:44
3	R1	10/02/2020	9:45	11:42
3	R2	11/02/2020	10:10	11:26
4	R1	12/03/2020	10:05	12:18
4	R2	13/03/2020	10:32	11:56

Fonte: autoria própria.

O RPA DJI *Matrice* 100 é uma plataforma totalmente moldável e pode ser utilizada por diferentes sensores e operações variadas (Figura 9A). Ele atinge uma altitude máxima de 170 m, chegando a 22 m/s e distância de transferência de 5 km. Destaca-se por sua estabilidade e autonomia de voo de 20 minutos, podendo ser duplicada com um slot adicional para bateria (DJI, 2020).

Acoplado ao RPA foram instalados uma câmera multiespectral *MicaSense RedEdge* e um sensor de irradiância (DLS) (Figura 9B). O sensor multiespectral *RedEdge* utilizado neste estudo coleta dados de cinco bandas espectrais simultaneamente: azul (475 nm, largura de banda de 20 nm), verde (560 nm, largura de banda de 20 nm), vermelho (668 nm, largura de banda de 10 nm), *rededge* (717 nm, 10 largura de banda de nm) e infravermelho próximo (840 nm, largura de banda de 40 nm), que foram correlacionas para a geração do modelo com maior acurácia (MICASENSE, 2020).

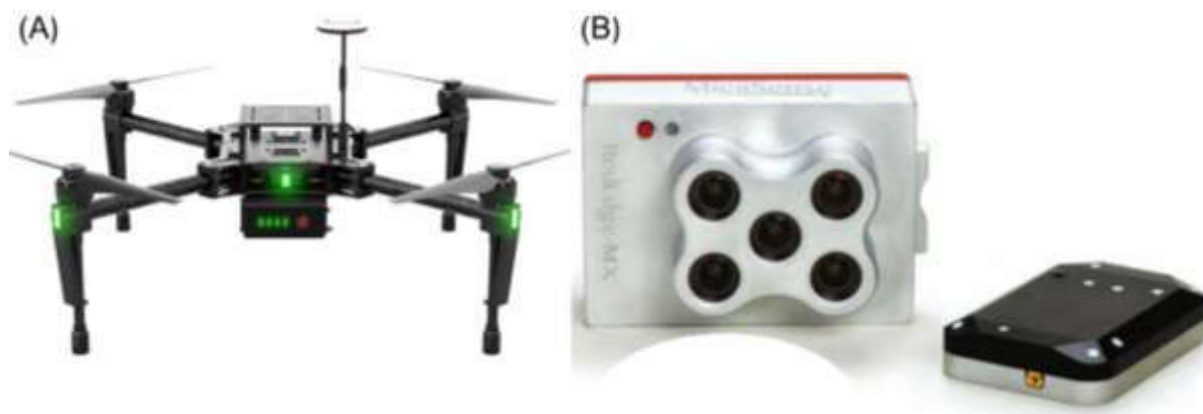


Figura 9 - RPA DJI Matrice 100 foi usado no estudo (A) e Sensor MicaSense RedEdge e DLS (B).
Fonte: (A) DJI, (2020), (B) Micasense, (2020).

O plano de voo seguiu as recomendações de Barbosa (2005), com horário de realização entre 9 e 14 horas, sempre se buscando o meio dia solar, quando o sol se encontra a pino, gerando menor sombreamento. Os voos foram realizados no sentido dos paralelos evitando manchas solares, com altitude de 120 metros e sobreposição lateral e frontal de 80% (Figura 10).

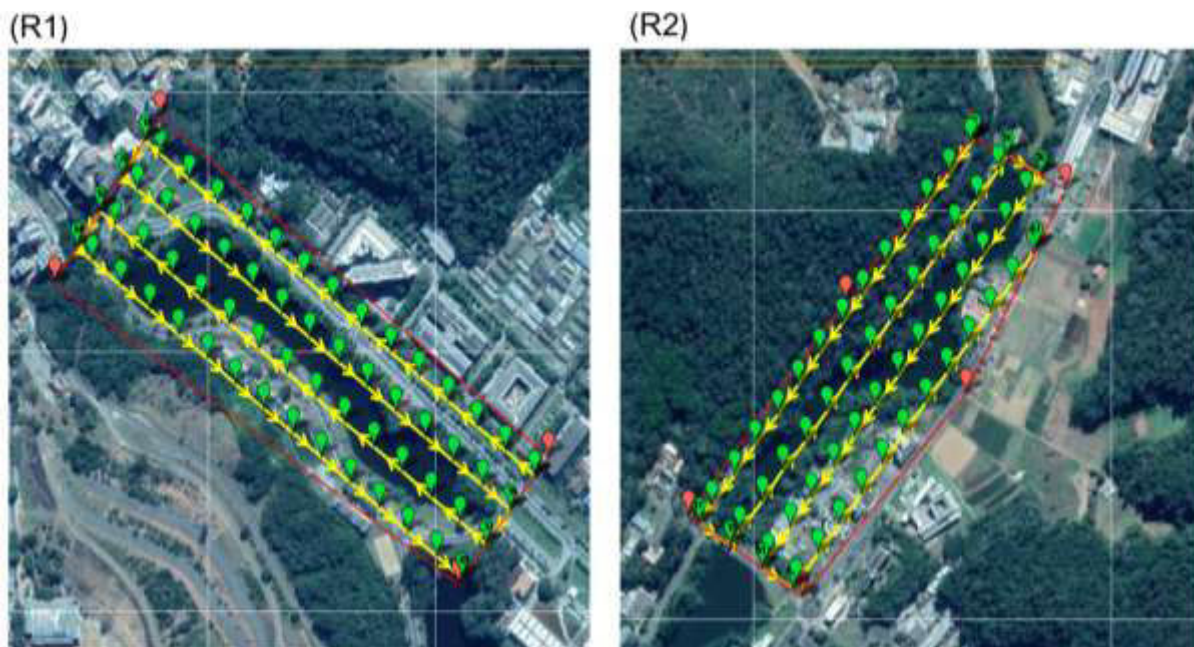


Figura 10 - Trajeto realizado pelo RPA no levantamento aéreo dos Reservatórios R1 e R2, representado pelas setas amarelas, e em balões verdes o momento de cada captura de imagem.
Fonte: autoria própria.

4.2 Correção Geométrica

Com o objetivo de garantir a acurácia do georreferenciamento das imagens, foram definidos dez locais para a coleta de pontos de controle no reservatório 1 e oito para o reservatório 2 (Figura 11A e B). Os pontos foram previamente coletados em locais fixos na paisagem (faixas de pedestre, meio fios, bocas de lobo), sendo usados os mesmo marcadores para todas as campanhas (Figura 11C). A ampla visualização dos marcos também foi observada, garantindo que eles não estivessem cobertos por árvores ou eventuais obstáculos. Para coleta dos pontos foi utilizando o receptor GNSS RTK TPS T10, que trabalha com todas as constelações de satélites, dez frequências e utiliza todos os métodos de conectividade disponíveis (Figura 11D).



Figura 11 – Distribuição dos pontos de controle no reservatório 1 (A) e no reservatório 2 (B), demarcação dos pontos de controle em alvos já existentes no local (C) e base do GNSS RTK TPS T10 instalada (D).

Fonte: autoria própria.

4.3 Calibração radiométrica

Calibração radiométrica é o processo de conversão do número digital (ND) para uma quantidade física a reflectância (IQBAL; LUCIEER; BARRY, 2018). Assim, antes e depois de cada voo foram realizadas imagens do painel de calibração de reflectância fornecido pela fabricante do sensor (Figura 12), afim de avaliar as diferentes condições de iluminação do ambiente.

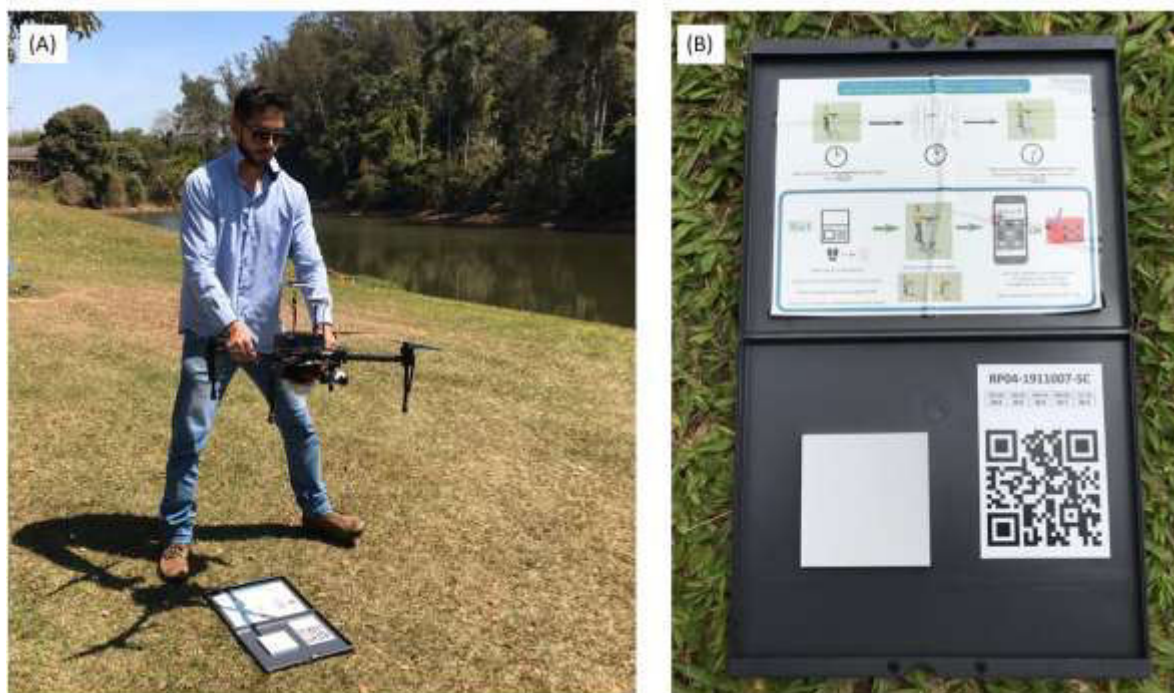


Figura 12 - Procedimento de calibração radiométrica das imagens em campo (A) e painel de calibração radiométrica (B).

Fonte: autoria própria.

4.4 Ortorretificação e mosaico das imagens aéreas

O processamento das imagens aéreas foi realizado no software *Agisoft Metashape* 1.6.1, onde inicialmente foram realizadas as correções geométricas provenientes da variação de movimentos do RPA e de inconsistências nas sobreposições entre imagens (SONG et al., 2016). Nesta etapa de pré-processamento foi aplicado o filtro *Image quality*, que identificou imagens com ruídos e excessos de contraste ou brilho, sendo que todas que possuíam qualidade inferior a 0,7 foram descartadas do processamento (AGISOFT, 2020).

Foram capturadas 504 imagens no R1 e 684 imagens no R2 em cada campanha, sendo essas imagens combinadas para a geração de um mosaico para cada reservatório. O mosaico é a técnica de junção de imagens que combina várias cenas em uma única imagem, cobrindo uma grande área (KIA; DOCKSTADER, 2013). No *Agisoft Metashape* o mosaico é realizado através de um algoritmo *Scale Invariant Feature Transform* (SIFT), com base na procura de pontos homólogos nas fotografias e pela aplicação de técnicas de *Structure from Motion* (SFM) e *Multiview Stereo* (MVS). Para garantir o georreferenciamento da imagem foram inseridos 10

pontos de controle nas bordas do reservatório. Como produto, se obteve um ortomosaico com *Ground Sample Distance* (GSD) de 7,35 cm.

Por fim, antes de realizar a plotagem do ortomosaico foi necessário converter a unidade de reflectância de cada banda, uma vez que ainda estavam com o valor de inteiro como o de entrada 16 bits (65.536 tons de cinza), representando 200% do valor real (Equação 8). Assim, o valor que refere a 100% do valor da reflectância equivale à 32.768 tons de cinza (MICASENSE, 2019). Esse procedimento foi realizado para cada banda individualmente com base na equação 8.

$$R_{Bi} = \frac{B_i}{32768} \quad (8)$$

Onde, R_{Bi} se refere a reflectância após a transformação e B_i é a banda a ser transformada.

4.5 Extração dos valores de reflectância

Após a geração dos ortomosaicos, as imagens foram inseridas no ArcMap 10.5, onde se iniciou um processo de refinamento e extração das reflectâncias. Em decorrência da movimentação superficial da água ocorre a reflexão especular de alguns *pixels* da superfície (CARVALHO; CABÚS, 2020). Assim, valores de reflectância tendem a superestimar a reflectância verdadeira. Por isso, todos os pixels com reflectância superior a 0,6 foram substituídos por valores *NoData*.

Para realizar a extração dos valores de reflectância das imagens multiespectrais, cinco cenários distintos foram considerados: valor pontual do pixel, onde a posição georreferenciada de cada estação de amostragem foi usada para extrair a reflectância; extração média do valor de reflectância de quadrados de 50, 100, 200 e 400 centímetros para cada lado, sendo criado nestes cenários um quadrado em torno da posição georreferenciada de cada estação de amostragem para extrair a reflectância média de cada uma das cinco bandas.

Além das bandas individuais foram geradas diversas razões de banda e índices espectrais com as cinco bandas para todos os tamanhos de amostras extraídos. Assim, foram geradas 85 variáveis para cada cenário descrito acima, com o propósito de gerar variáveis com grande correlação com o SST.

Os valores de reflectância de todos os cenários foram inicialmente inseridos como covariáveis para a criação do modelo de aprendizado de máquina e, dessa forma, o algoritmo selecionou o conjunto ideal de covariáveis para cada modelo. Todo esse procedimento foi realizado em linguagem R no *software* R Studio (R CORE TEAM, 2020).

4.6 Campanhas de coleta das amostras de água

As coletas das amostras de água em campo foram realizadas previamente aos voos do RPA, representando as mesmas condições nas duas plataformas. Assim, foram demarcados 30 pontos aleatórios em cada reservatório, com densidade amostral de 9,23 amostras/ha, valor superior aos utilizados por Bonansea et al. (2015b); Su; Chou (2015) e Zhao et al. (2011). Este valor foi escolhido para permitir melhor análise espacial dos dados, de forma a ser estatisticamente representativo.

Foi utilizado o GNSS RTK TPS T10 para realizar a demarcação das coordenadas em todos os locais de coleta das amostras. As coletas foram realizadas com um béquer adaptado com um bastão graduado, permitindo coletar água subsuperficial (em profundidade de 30 cm abaixo da superfície). Em cada ponto foram coletados aproximadamente 2 litros de água, que foram armazenadas em garrafas pet devidamente etiquetadas.

4.6.1 Análise das amostras

Todas as garrafas foram encaminhadas imediatamente para a Divisão de Água e Esgoto da Universidade Federal de Viçosa, onde foram acondicionadas em ambiente com temperatura inferior a 4°C. As análises para a determinação dos sólidos suspensos totais foram realizadas conforme metodologia descrita por ALPHA, AWWA e WEF (2017). Inicialmente, foi realizado o preparo das membranas de fibra de vidro 47 mm x 1,2 µm; as quais foram colocadas na estufa por 1 hora a uma temperatura de 105°C e em seguida levadas para o dessecador por 24 horas a fim de eliminar a umidade. Posteriormente à secagem, quantificou-se o peso das membranas em balança analítica para obtenção do peso inicial (P_i) (Figura 13).

O processo de filtração foi realizado em uma bomba de vácuo de 1 HP. Após a realização da filtração, as membranas foram colocadas novamente na estufa por

1 hora a 105°C e levadas para o dessecador por 24 horas, sendo pesadas na mesma balança e obtendo-se o peso final (P_f) (Figura 13B). Desta forma, foi determinada a concentração de SST em mg/L como demonstrado na equação 9. Para cada ponto de coleta foram filtrados entre 200 e 500 mL em função dos sedimentos presentes na coluna d'água. As amostras foram feitas em duplicatas, o que resultou em variáveis quantificadas a partir da média entre estas análises.

$$SST = \left[\frac{pf - pi}{V} \right] \times 1000 \quad (9)$$

Onde: SST - Sólidos Suspensos Totais; pf - Peso Final (g); pi - Peso Inicial (g); V - Volume (L).

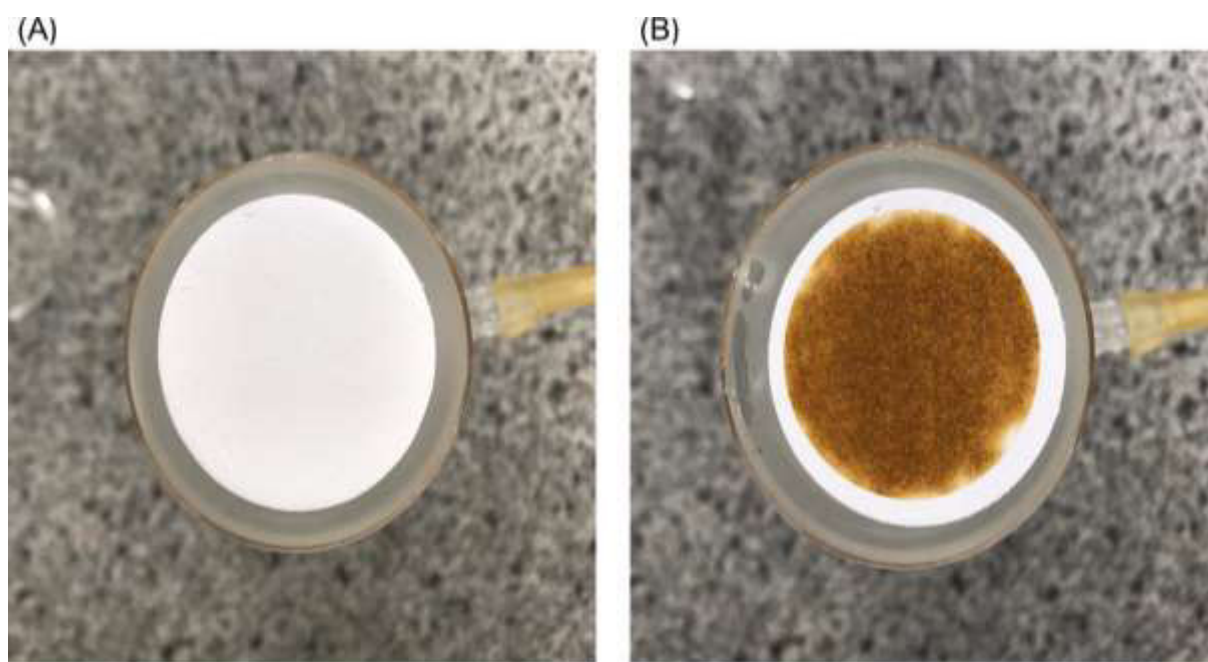


Figura 13 - Membranas de fibra de vidro antes (A) e após (B) o processo de filtração do SST.
Fonte: autoria própria.

4.7 Levantamento batimétrico dos reservatórios

O levantamento batimétrico dos reservatórios de acumulação de água foi realizado por Carmo e Ferreira (2012). Neste levantamento foi utilizado um Ecobatímetro *Hydrobox*, feixe simples, com precisão nominal de 0,5%. Para nortear o levantamento, um GNSS RTK TPS T10 foi utilizado.

4.8 Modelagem de SST

Visando a obtenção de estimativas confiáveis de SST foram aplicadas três diferentes abordagens para a otimização dos procedimentos de estimativa (Tabela 3). As abordagens variaram de acordo com os dados de entrada, com as seguintes premissas sendo estabelecidas: A1 – Treinamento no reservatório 1 e validação no reservatório 2; treinamento no reservatório 2 e validação no reservatório 1; A2 – Treinamento em 3 campanhas e validação em 1 campanha (as campanhas foram definidas aleatoriamente, sendo esse processo repetido até se realizar todas as combinações possíveis); A3 – agrupou-se os dados do reservatório 1 com os dados do reservatório 2 e dividiu-se 75% dos dados para treinamento e 25% para validação aleatoriamente, repetindo este procedimento por 100 vezes.

Tabela 3 - Abordagens utilizadas na modelagem de SST

Abordagens	Variáveis de entrada	
	Treinamento	Validação
A1	Dados R1 Dados R2	Dados R2 Dados R1
A2	Dados de três Campanhas	Dados de uma Campanha
A3	75%	25%

Fonte: autoria própria.

Na Figura 14 são apresentados os procedimentos metodológicos aplicados para a modelagem de SST nos reservatórios 1 e 2.

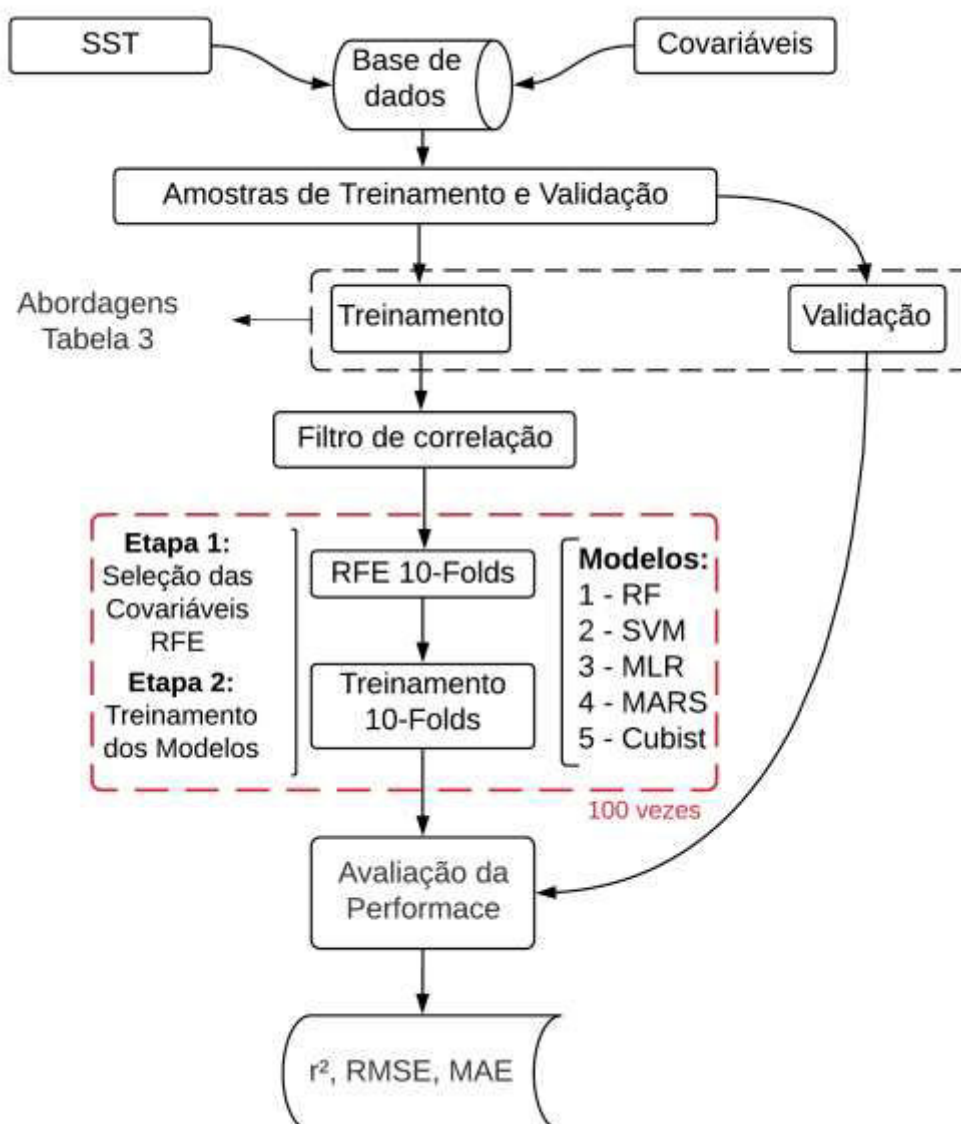


Figura 14 - Esquema metodológico dos processos aplicados para a estimativa de SST nos reservatórios 1 e 2.

Fonte: autoria própria.

4.8.1 Seleção das covariáveis

A seleção de variáveis é frequentemente usada para reduzir a dimensionalidade do espaço de recursos e remover recursos ruidosos e redundantes (MUÑOZ-ROMERO et al., 2020; REUNANEN, 2003; ZHANG et al., 2020). O objetivo é selecionar um pequeno subconjunto de recursos originais que minimize a redundância e maximize a relevância, reduz o número de componentes do modelo, utilizando somente as mais importantes para prever a variável dependente em análise. Assim, foi aplicado o coeficiente de correlação de Pearson

(LEE RODGERS; NICEWANDER, 1988) eliminando todas as covariáveis com correlação superior a 95%.

Posteriormente, o algoritmo *Recursive Feature Elimination* (RFE) do pacote *Caret* foi implementado. Este é um algoritmo de seleção reversa que elimina iterativamente os preditores menos promissores do modelo, com base em uma medida inicial de importância do preditor (KUHN, 2020). Tem sido amplamente utilizado para o mapeamento ambiental (GHOSH; JOSHI, 2014; GOMES et al., 2019; MEYER et al., 2017, 2018; STEVENS et al., 2013), sendo aplicado nas covariáveis bandas espectrais, razões entre bandas, índices de bandas, batimetria, coordenadas e meses para a seleção do grupo mais representativo para a estimativa de SST no presente trabalho. O detalhamento das covariáveis inseridas nos modelos estão detalhadas no Apêndice A.

4.8.2 Seleção dos modelos de aprendizado de máquina

Para a estimativa do SST foram escolhidos cinco diferentes modelos de aprendizado de máquina: *Random Forest* (RF), *Support Vector Machines Radial Sigma* (SVM-RS), *Multivariate Adaptive Regression Spline* (MARS), *Multiple Linear Regression* (MLR) e *Cubist*. Esses métodos foram selecionados para abranger uma variedade de diferentes modelos linear, não linear, baseados em regras e em árvore de regressão. A principal diferença entre os modelos é a forma como realizam o gerenciamento dos dados em seu conjunto de treinamento. Assim, uma variável importante para um dos modelos pode não ser representativa para outro. Cabe salientar que, para avaliar estritamente cada modelo, foi aplicado o RFE gerando seu conjunto ideal de covariáveis. Isto significa que existem cinco conjuntos diferentes de variáveis para realizar a estimativa de SST. Para mais informações sobre eles, consulte a publicação de Kuhn e Johnson (2013); Murphy (2013).

4.8.3 Avaliação dos modelos de aprendizado de máquina

Para selecionar o melhor modelo, cada um deles foi aplicado separadamente usando seu próprio subconjunto ideal de covariáveis selecionadas pela análise RFE. Os dados foram modelados seguindo as abordagens da Tabela 3 e a precisão foi expressa pelos índices estatísticos: coeficiente de determinação - r^2 (Equação 10), raiz do erro quadrático médio – RMSE (Equação 11), erro absoluto médio –

MAE, (Equação 12). A abordagem A2 foi repetida por quatro vezes e a A3 por 100 vezes com o seu próprio subconjunto de variáveis e comparadas com os valores médios dos parâmetros de precisão (r^2 , RMSE e MAE). A adoção de várias repetições é importante para determinar a variabilidade da previsão, pois diferentes grupos de conjuntos de dados de treinamento e validação podem gerar diferentes resultados de predição (KUHN e JOHNSON, 2013).

$$R^2 = \frac{\sum_{i=1}^n (P_i - Om)^2}{\sum_{i=1}^n (Om_i - \overline{Om})^2} \quad (10)$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad (11)$$

$$MAE = \frac{\sum_{i=1}^N |P_i - O_i|}{n} \quad (12)$$

4.9 Mapa de estimativas de SST dos Reservatórios

Para construir os mapas de estimativas de SST dos reservatórios foram selecionados, dentre os cinco modelos de aprendizado de máquina, o que apresentou o melhor desempenho na comparação dos valores preditos com os observados nas campanhas de campo. Este modelo foi executado por 100 vezes, gerando 100 mapas de estimativa de SST, sendo que o mapa final foi gerado com o valor médio resultante de cada pixel.

5 RESULTADOS E DISCUSSÃO

5.1 Amostragem de SST

No total foram coletadas 120 amostras de água para cada reservatório, em quatro campanhas de diferentes condições. A concentração de SST no Reservatório 1 variou entre 3,75 e 58,25 mg/L e no Reservatório 2 foi de 2,67 a 29,75 mg/L. Analisando as características construtivas de cada reservatório observou-se que o R1 possui diversas entradas de água pluvial em toda sua estrutura, fato que não acontece no R2. Allen et al. (2017) apontam que, em áreas urbanas impermeáveis, o fluxo de água da chuva no solo causa a coleta de poluentes e sedimentos, que são transportados para as hidrografias mais próximas.

Na Figura 15 apresenta-se os resultados do boxplot das concentrações de SST nas quatro campanhas realizadas nos dois reservatórios.

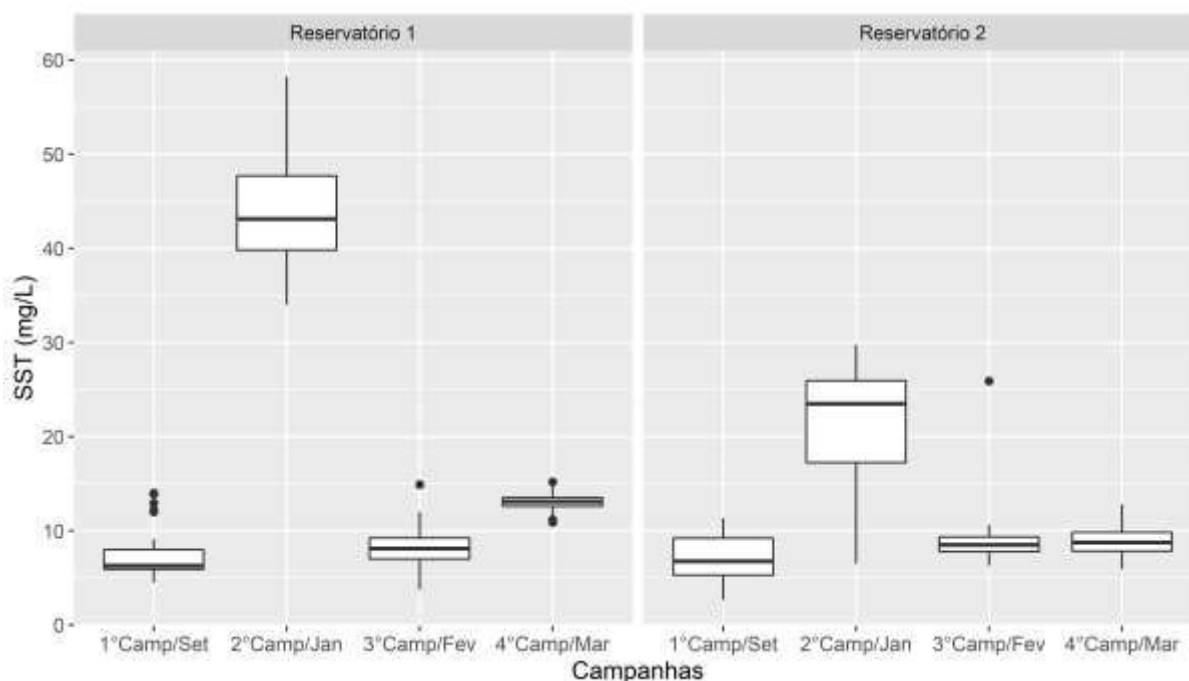


Figura 15 – Boxplot das amostras de SST nas campanhas de setembro/2019, janeiro/2020, fevereiro/2020 e março/2020 dos dois reservatórios.

Fonte: autoria própria.

Quando comparadas as amostras de cada campanha, percebe-se uma distribuição bem regular entre o valor mínimo e os intervalos interquartis. Entretanto, quando se avaliam os valores entre as campanhas percebe-se uma maior

heterogeneidade, com destaque para a 2ª campanha, realizada em janeiro, onde se contabilizou as maiores concentrações de SST nos dois reservatórios. Esses valores elevados podem ser justificados pois, apesar de não chover nos dias das amostragens, houve eventos de chuva na semana que antecedeu as campanhas, (208,6mm) de acordo com os dados da estação A510, localizada na cidade de Viçosa (INMET, 2020). Desta forma, houve carreamento dos sedimentos da bacia hidrográfica até os reservatórios.

Na Figura 16 observa-se os diferentes comportamentos dos espectros amostrados nas campanhas de janeiro/2020 e setembro/2019. A assinatura espectral das amostras corrobora com os dados de Barbosa, Novo e Martins (2019), que realizaram experimentos com diferentes concentrações de SST com o intuito de se identificar a assinatura espectral das soluções. Já Barsi, Markham e Pedelty (2011); Doxaran et al. (2002) que evidenciaram que as águas turvas têm propriedades ópticas diferentes das águas limpas que resultam em alterações na resposta espectral das leituras de sólidos suspensos totais.

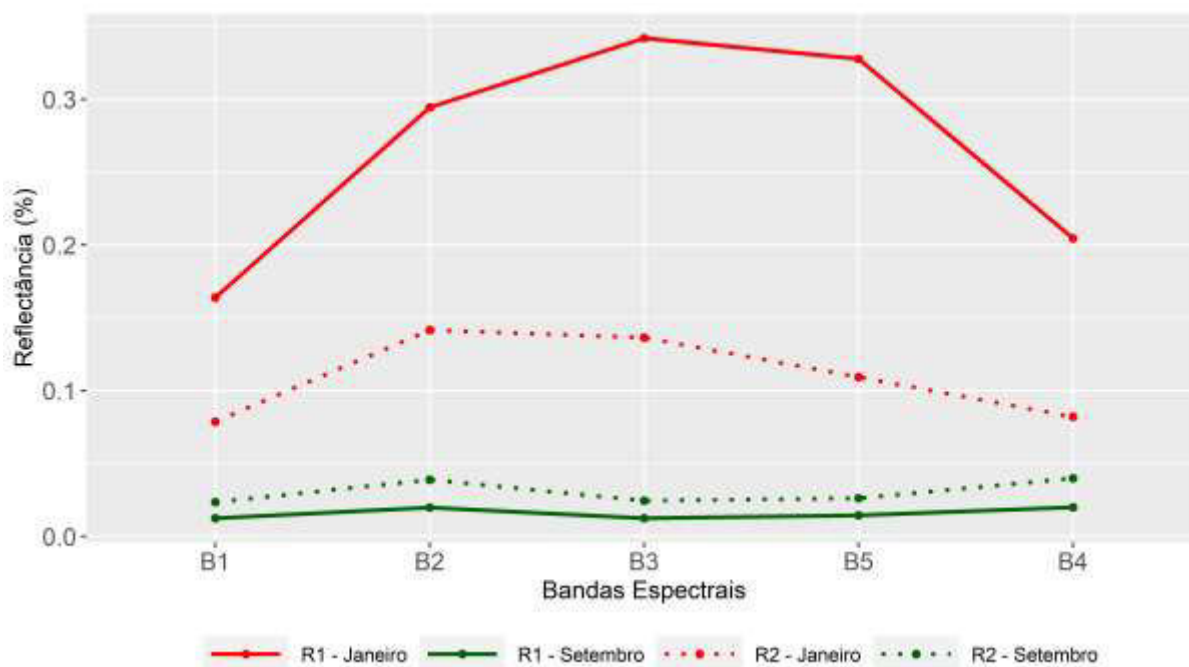


Figura 16 - Comportamento radiométrico médio das amostras coletadas com sensor multiespectral nas campanhas de janeiro/2020 e setembro/2019.

Fonte: autoria própria.

Verifica-se na Figura 16 que a reflectância espectral das amostras do mês de setembro/2019, nos dois reservatórios, é baixa e moldada principalmente pelas

propriedades ópticas das moléculas de água pura, que produzem um aumento significativo no espalhamento em direção aos comprimentos de onda mais curtos (B1) e de absorção em direção aos comprimentos de onda mais longos (B4). Já nas campanhas de janeiro, observa-se aumento das concentrações de SST, provocando aumento da reflectância em toda a faixa espectral estudada (BARBOSA; NOVO; MARTINS, 2019; CHEN; CURRAN; HANSOM, 1992). Nota-se valores mínimos de reflectância na B1 (faixas de absorção de radiação, em torno de 450nm), tendo uma forte curva de espalhamento em direção a B3, devido às propriedades espectrais conjuntas dos sedimentos em suspensão e da matéria orgânica dissolvida. Depois a marca de evidente declínio até a B4, justificado pela rotação da molécula de água (RUDORFF et al., 2007).

5.2 Seleção das covariáveis

Com a aplicação do filtro de correlação de Pearson houve uma redução de 430 covariáveis para 60. Posteriormente foi realizada a aplicação do RFE, observando-se a seleção de um número distintos de covariáveis importantes para cada um dos modelos de aprendizado de máquina utilizado. Na Figura 17 apresenta-se o desempenho dos cinco modelos no processo de seleção de variáveis em cada uma das três abordagens utilizadas (A1, A2 e A3). Verifica-se que, em média, 16 variáveis tiveram desempenho igual ou superior ao conjunto total de variáveis. Muñoz-Romero et al. (2020) conseguiram otimizar significativamente o desempenho do seu modelo aplicando filtro de correlação de Pearson. Stevens et al. (2013) reduziram o número de variáveis ambientais de seu modelo aplicando o RFE, diminuindo consideravelmente o tempo de processamento dos modelos.

Os modelos *Cubist*, RF e MARS tiveram desempenhos similares, utilizando a menor quantidade de variáveis explicativas para compor o modelo. Já os modelos MLR e SVM-RS necessitaram de uma quantidade maior de variáveis para conseguir bons resultados. O modelo MLR apresenta significativa sensibilidade com o aumento da quantidade de variáveis, apresentando um destaque negativo em comparação aos outros modelos.

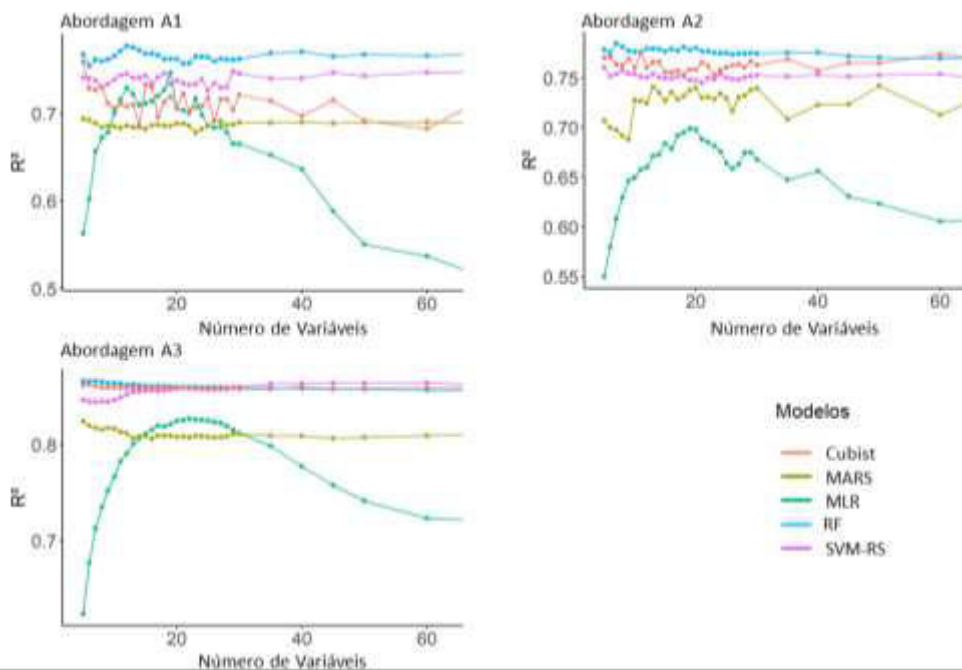


Figura 17 - Desempenho dos modelos de aprendizado de máquina no processo de seleção de variáveis para cada uma das três abordagens avaliadas.

Fonte: autoria própria.

A Figura 18 apresenta a dispersão dos dados observados com os dados preditos em uma reta 1:1. Ao avaliar as três abordagens se identifica uma grande discrepância nos valores da abordagem A1 em comparação com a A3. Isso se deve à dificuldade dos modelos na extrapolação do intervalo de dados para o qual foi treinado, tendo em vista que o intervalo de SST do R2 é 2,67 a 29,75 mg/L e do R1 é 3,75 e 58,25 mg/L. Na abordagem A2, onde se realizou o treinamento com 3 campanhas completas e validação com 1 campanha completa, observa-se o mesmo comportamento. Quando se realiza o treinamento apenas com as campanhas de baixa concentração de SST (setembro, fevereiro e março) e validação com as campanhas de alta concentração (janeiro) os modelos têm baixo desempenho; entretanto, quando se realiza qualquer outra combinação incluindo o mês de janeiro no treinamento se consegue bons resultados. Hengl et al. (2018); Zhang, Li e Deng (2020) tiveram experiência parecidas e relataram que os modelos treinados ficam aquém da extrapolação de dados para longe do espaço do conjunto de dados de treinamento, embora exibam desempenho de interpolação excelente em alguns casos.

Diante disso, a abordagem A3, onde se realizou o treinamento com 75% dos dados e validação com 25% dos dados aleatórios, repetidos por 100 vezes, se mostrou a mais adequada no presente trabalho e foi a que melhor se ajustou na reta

1:1, apresentando um ótimo indício de que os dados preditos não subestimam nem superestimam o SST.

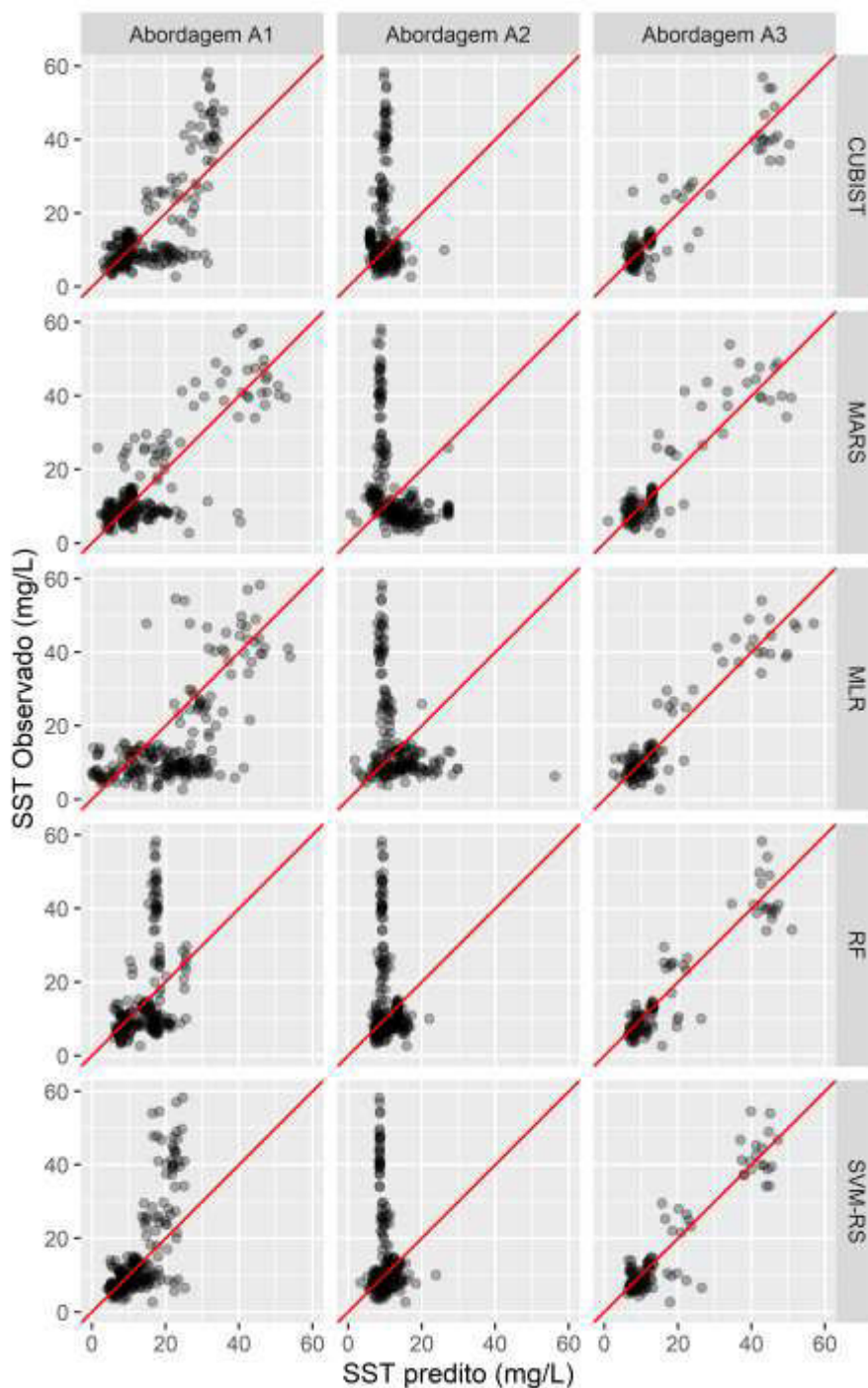


Figura 18 - Comparação entre os valores observados e estimados de SST nos modelos RF, SVM-RS, MLR e MARS e Cubist nas três, aplicando diferentes abordagens.

Fonte: autoria própria.

Ao analisar a Tabela 4, que apresenta o desempenho dos modelos RF, SVM-RS, MARS, MLR e *Cubist* na estimativa de SST para a abordagem A3, observa-se que por se tratar de modelos robustos todos os modelos tiveram um bom desempenho, entretanto, os modelos RF, SVM-RS se destacaram. Ao realizar as análises dos índices estatísticos identificou-se que o modelo SVM-RS foi superior para a previsão das concentrações de SST, com maior valor de r^2 e menores valores de RMSE e MAE, o que corrobora com estudos de Zhang, Huang e Wang (2020). Estes autores conseguiram excelentes estimativas de parâmetros de água utilizando o algoritmo SVM. Já Hengl *et al.*, (2018); Zhang, Li e Deng (2020) identificaram que algoritmos de árvore de decisão apresentam limitação na modelagem de conjuntos com menor número de dados, com pouca aleatoriedade, e quando precisa realizar a extrapolação dos dados para valores além intervalo de treinamento.

Com as estatísticas de predição do conjunto treinamento e validação é possível identificar que o modelo SVM-RS não apresenta qualquer *overfitting*. O RMSE e MAE corroboram com essa afirmação, apresentando pequenas diferenças entre si e pouca variação entre os valores de treinamento e validação. O modelo MLR apresentou valores satisfatórios, o que poderia ser uma opção interessante, pois pode ser facilmente aplicado através de uma equação linear, desde que os valores estejam dentro do intervalo desta pesquisa.

Tabela 4 – Desempenho dos modelos RF, SVM-RS, MARS, MLR e Cubist na estimativa de SST nas três diferentes abordagens

Modelos	Abordagem	Treinamento			Validação		
		r^2	RMSE	MAE	r^2	RMSE	MAE
RF	A3	0.875	4.292	2.872	0.865	4.696	3.080
SVM-RS	A3	0.870	4.342	2.975	0.869	4.559	3.021
MLR	A3	0.851	4.784	3.424	0.840	5.064	3.555
MARS	A3	0.865	4.483	3.179	0.843	4.972	3.409
Cubist	A3	0.871	4.428	2.948	0.850	4.896	3.151

Fonte: autoria própria.

Após a comprovação de que a abordagem A3 e o algoritmo SVM-RS apresentaram o melhor desempenho na estimativa de SST, foi realizada uma análise de importância de variáveis (Figura 19). Observa-se através desta análise que as razões de banda se destacam como covariáveis mais importantes na estimativa do SST comparadas às bandas simples, corroborando com os resultados

encontrados por Sestini (1999), que identificou que razões entre bandas permite discriminar sutis diferenças existentes no comportamento espectral de diferentes alvos, pois em bandas simples apenas diferenças discrepantes são identificadas. Doxaran et al. (2002), realizando estudos para quantificar material particulado suspenso em água, notaram que a reflectância de uma banda única pode ser fortemente influenciada por variações nas características das partículas, características do sensor e condições de luz, que podem ser melhoradas usando razões de bandas e índices espectrais.

As razões de bandas envolvendo B1 (Azul), B4 (NIR) e B5 (rededge) destacaram das demais no processo de seleção de variáveis, apresentando alta importância na construção do modelo. Analisando as características do solo da região, pode se observar que se trata de latossolos vermelho-amarelos (UFV et al., 2010), sendo um indicio da predominância de bandas espectrais entre 700 e 850nm. Sáens, Paez e Arango (2015) obtiveram bons resultados, com $r^2 = 0,87$, utilizando as bandas 1 (Azul) e banda 4 (NIR) na identificação de SST em microbacia na Colômbia. Gholizadeh, Assefa e Reddi (2016); Matthews (2011) encontraram bons resultados com razões de banda entre 670 a 710nm. É importante ressaltar que se deve analisar as características do local de estudo, como solo e vegetação, pois estes têm um papel primordial na seleção das bandas em diferentes locais. As covariáveis com tamanhos amostrais de 400cm, 200cm e pontual foram as de maior importância, estando presentes em 70% das variáveis importantes do modelo (Figura 19).

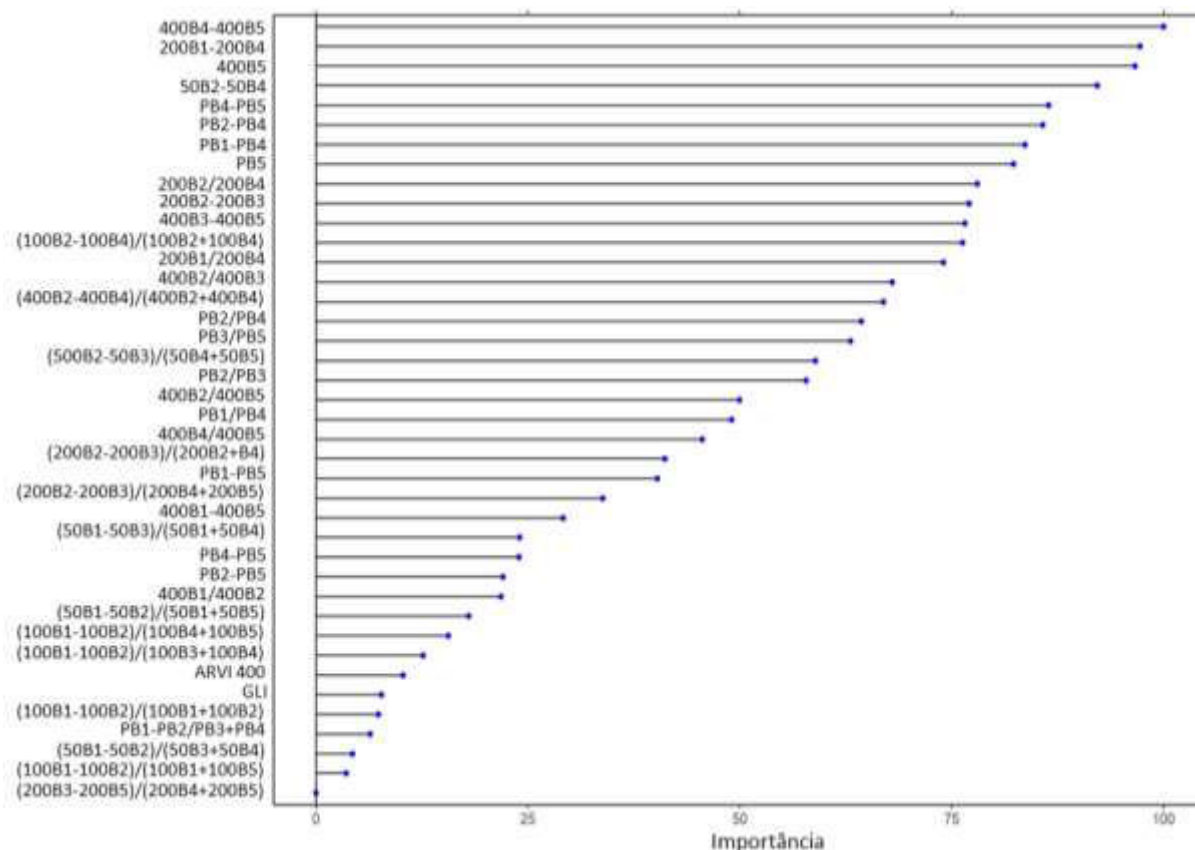


Figura 19 – Importância das covariáveis na estimativa de SST utilizando o modelo RFE-SVM-RS. (P: valor pontual, 50: 50 cm, 100: 100 cm, 200: 200 cm, 400: 400 cm)
Fonte: autoria própria.

Na Figura 20 pode-se observar a distribuição espacial da concentração de SST predita pelo modelo SVM-RS das diferentes campanhas no reservatório 1. Através dos mapas observa-se que apenas a campanha de janeiro destoa das demais. Os valores mais altos em todas as campanhas estão localizados nas extremidades sul, onde se localizam um canal com entrada de água, e a norte, onde há um extravaso, sendo possível identificar um fluxo central de deslocamento do sedimento. Cabe ressaltar que as bordas do reservatório apresentam pequenas contaminações em decorrência de pixels com composição mista e de áreas assoreadas.

A análise temporal nos reservatórios foi de extrema relevância, pois demonstrou a aplicabilidade do modelo em diferentes condições, enfatizando sua capacidade de generalização espaço-temporal, o que corrobora com as análise de Guo et al. (2015). Estes pesquisadores conseguiram ótimos valores na predição de parâmetros ópticos e não ópticos de qualidade de água com sensor multiespectral acoplado a RPA. No entanto, aplicaram modelos lineares para desenvolvimento dos modelos. Zhang, Huang e Wang (2020) ressaltam a importância de se realizar coleta

de dados em várias condições, de modo a deixar o modelo mais sensível às variações ambientais. Estes autores utilizaram dados de sensoriamento remoto e o algoritmo SVM para a predição de parâmetros de qualidade da água.

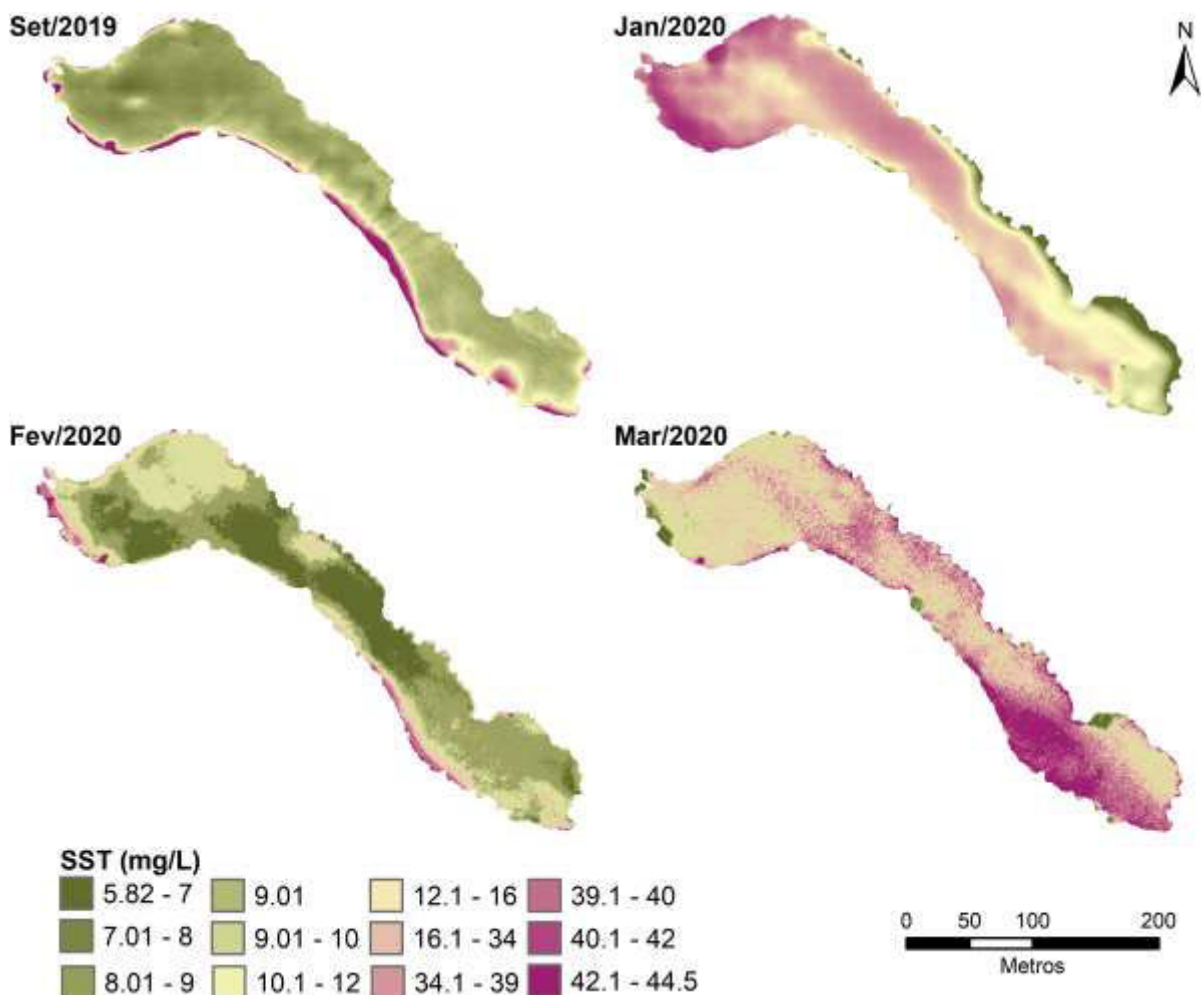


Figura 20 - Predição da concentração de SST pelo modelo SVM-RS para o reservatório 1 nas quatro campanhas realizadas.
Fonte: autoria própria.

Na Figura 21 é possível avaliar as predições da concentração de SST do modelo SVM-RS para o reservatório 2. Este reservatório tem um menor intervalo entre os valores de SST entre as campanhas, com os maiores valores encontrados na porção sul do reservatório, onde possui um canal de entrada de água que causa grande revolvimento de solo e água colocando as partículas em suspensão. A campanha do mês de setembro apresenta uma lacuna na parte sul, por se tratar de um trecho assoreado, com o baixo nível da água ficou exposto no período de estiagem (setembro).

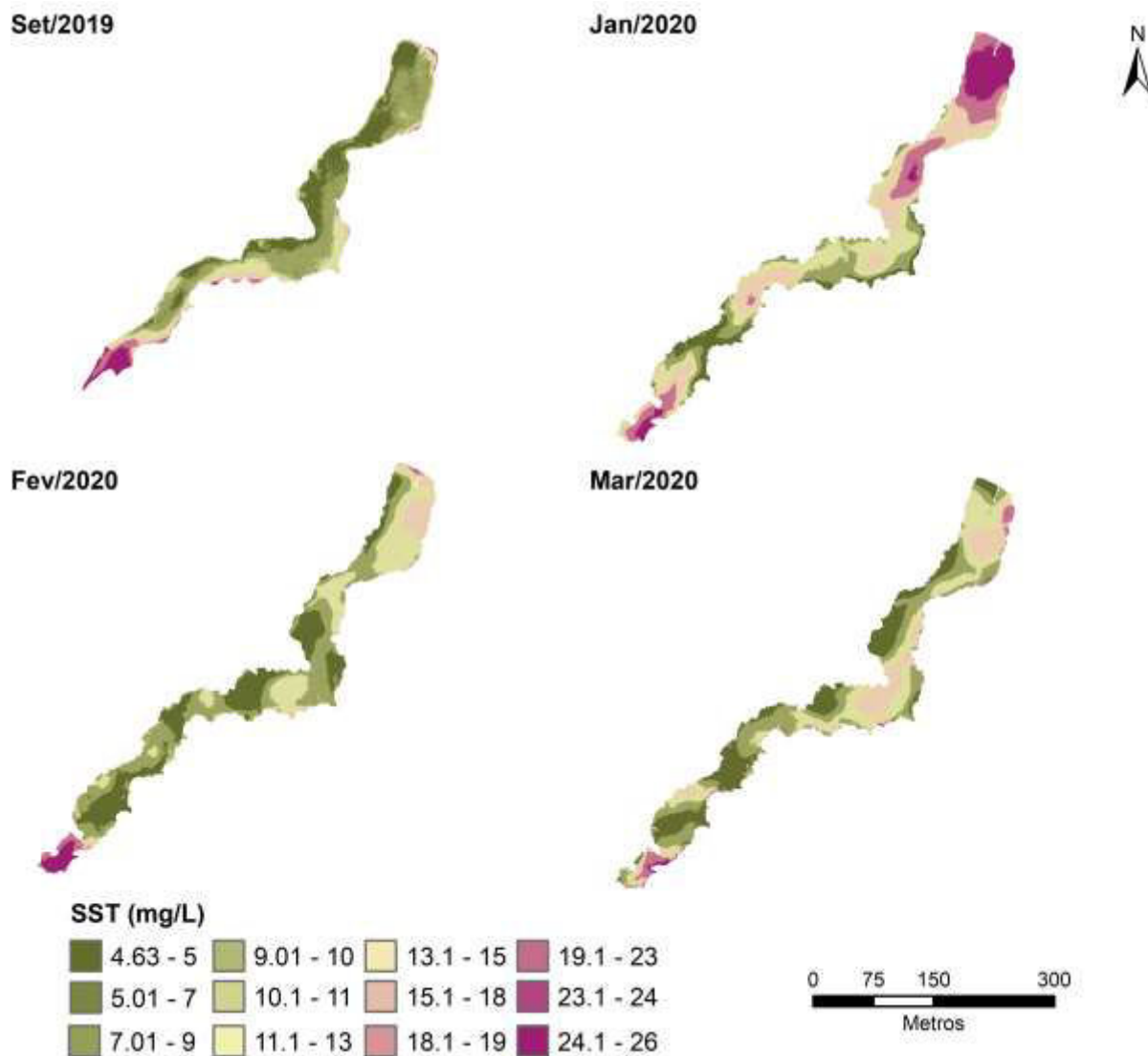


Figura 21 - Predição da concentração de SST pelo modelo SVM-RS para o reservatório 2 nas quatro campanhas realizadas.
Fonte: autoria própria.

As análises multitemporais foram de grande importância para este estudo, não apenas para a validação dos modelos, mas também porque demonstra a variação da concentração de SST em diferentes condições do ano. Bilotta e Brazier (2008) realizaram uma revisão minuciosa sobre o comportamento do sólido suspenso e deixaram clara a importância do monitoramento espaço-temporal contínuo para a gestão adequada dos recursos hídricos.

Por todas as informações discutidas, pode-se inferir que o uso de sensor multiespectral acoplado a RPA oferece benefícios adicionais comparativamente à metodologia tradicional de coleta de amostras pontuais de qualidade da água para fins de avaliação de SST. Sugere-se para futuros estudos, a realização de mais

campanhas no espaço e redução da quantidade de amostras no tempo, assim, obtendo maior diversificação das condições temporais do modelo.

6 CONCLUSÕES

A metodologia desenvolvida para a estimativa de SST em reservatórios, com base em dados de sensor multiespectral acoplado em Aeronave Remotamente Pilotada (RPA), apresentou-se adequada, obtendo valores acurados dos índices estatísticos na comparação entre valores preditos e observados, com potencial para ser usada para o monitoramento espaço-temporal do SST em reservatórios de acumulação de água.

O modelo de aprendizado de máquina SVM-RS (*Support Vector Machines Radial Sigma*) apresentou os melhores resultados na predição de SST, indicando que pode ser utilizado para quantificar remotamente o SST com elevado grau de confiabilidade.

Foi identificada, nas análises espectrais, forte relação dos espectros coletados com sensor multiespectral com as amostras coletadas in loco, possibilitando a avaliação da influência das condições ambientais nas diferentes condições de estimativa de SST. As bandas 1, 4 e 5 são extremamente importantes para estimativas mais acuradas e as razões de banda são mais robustas na predição de SST comparativamente às bandas simples.

A geração de mapas da concentração de SST dos reservatórios permitiu uma análise espaço-temporal, o que é primordial para o monitoramento e gestão adequada dos recursos hídricos.

REFERÊNCIAS

- ADAMS, A.; STERLING, L. **AI '92** WORLD SCIENTIFIC, , dez. 1992. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/9789814536271>>
- AGISOFT. **Agisoft Metashape User Manual: Professional Edition, Version 1.6** Agisoft, 2020. Disponível em: <https://www.agisoft.com/pdf/metashape-pro_1_5_en.pdf>
- ALLEN, D. et al. Multiple rainfall event pollution transport by sustainable drainage systems: the fate of fine sediment pollution. **International Journal of Environmental Science and Technology**, v. 14, n. 3, p. 639–652, 2017.
- ALPHA; AWWA; WEF. **Standard Methods for the Examination of Water and Wastewater** WASHINGTON - DC, 2017.
- ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.
- ANA. **Agência Nacional de Águas. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas** Brasília - DF, 2017.
- ARANGO, J. G.; NAIRN, R. W. Prediction of Optical and Non-Optical Water Quality Parameters in Oligotrophic and Eutrophic Aquatic Systems Using a Small Unmanned Aerial System. **Drones**, v. 4, n. 1, p. 1, 24 dez. 2019.
- ASNER, G. P. Cloud cover in Landsat observations of the Brazilian Amazon. **International Journal of Remote Sensing**, v. 1161, 2010.
- BARBOSA, C. C. F. **Sensoriamento remoto da dinâmica da circulação da água do sistema planície de Curuai/Rio Amazonas** Instituto Nacional de Pesquisas Espaciais (INPE), 2005.
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos**. 1° ed. São José dos Campos: INPE, 2019.
- BARSI, J. A.; MARKHAM, B. L.; PEDELTY, J. A. The operational land imager: spectral response and spectral uniformity. **Earth Observing Systems XVI**, v. 8153, n. September 2011, p. 81530G, 2011.
- BERKMAN, J. A. H.; CANOVA, M. G. **Algal biomass indicators** U.S. Geological Survey Reston, VA, 2007. Disponível em: <<http://pubs.water.usgs.gov/twri9A>>

- BEYCA, O. F. et al. Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. **Energy Economics**, v. 80, p. 937–949, 2019.
- BILOTTA, G. S.; BRAZIER, R. E. Understanding the influence of suspended solids on water quality and aquatic biota. **Sciencedirect**, v. 42, p. 2849–2861, 2008.
- BONANSEA, M. et al. Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina). **Remote Sensing of Environment**, v. 158, p. 28–41, 2015a.
- BONANSEA, M. et al. Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina). **Remote Sensing of Environment**, v. 158, p. 28–41, 2015b.
- BREIMAN, L. Random forests. **Machine learning**, v. 45.1, p. 27, 2001.
- CARMO, E. J. DO; FERREIRA, Í. O. **Mapeamento dos relevos submersos das lagoas da UFVViçosa - MG**, 2012.
- CARVALHO, M. L. S.; CABÚS, R. C. Eficiência da luz solar refletida e desempenho de dispositivos de sombreamento. **Ambiente Construído**, v. 20, n. 2, p. 191–209, 2020.
- CHEN, H. Monitoring tropical billabong water turbidity using Remotely Piloted Aircraft System (RPAS) derived imagery. n. October, p. 70, 2016.
- CHEN, Z.; CURRAN, P. J.; HANSOM, J. D. Derivative reflectance spectroscopy to estimate suspended sediment concentration. **Remote Sensing of Environment**, v. 40, n. 1, p. 67–77, 1992.
- DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing** (P. V. U. Amsterdam, Ed.) **Proefschrift Vrije Universiteit Amsterdam**, 2001.
- DJI. **DJI Matrice 100 USER MANUAL**.
- DOUGHTY, C. L.; CAVANAUGH, K. C. Mapping coastal wetland biomass from high resolution unmanned aerial vehicle (UAV) imagery. **Remote Sensing**, v. 11, n. 5, 2019.
- DOXARAN, D. et al. Spectral signature of highly turbid waters Application with SPOT data to quantify suspended particulate matter concentrations. **Remote Sensing of Environment**, v. 81, n. 1, p. 149–161, 2002.

- DUAN, W. et al. Spatiotemporal evaluation of water quality incidents in Japan between 1996 and 2007. **Chemosphere**, v. 93, n. 6, p. 946–953, 2013a.
- DUAN, W. et al. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River , Japan , 1985 to 2010. **Science of the Total Environment**, The, v. 461–462, p. 499–508, 1 set. 2013b.
- EMBRAPA, E. B. DE P. A. Sumula da X Reunião Técnica de Levantamentos de Solos (SNLCS, Série Miscelânea, 1). **Serviço Nacional de Levantamento e Conservação de Solos**, 1979.
- EMBRAPA, E. B. DE P. A. **Sistema brasileiro de classificação de solos** (EMBRAPA, Ed.)Rio de Janeiro, 2006.
- ENGMAN, E. T.; GURNEY, R. J. **Remote sensing in hydrology**. Londres: Chapman and Hall Ltd, 1991.
- FILGUEIRAS, R. et al. Crop NDVI monitoring based on sentinel 1. **Remote Sensing**, v. 11, n. 12, 2019.
- FLORENZANO, T. G. **Imagens de satélite para estudos ambientais**São Paulo - SPOficina de textos, , 2002.
- FRIEDMAN, J. H. Multivariate Adaptive Regression Splines. **the Annals of Statistics**, v. 19, n. 1, p. 1–67, 1991.
- GHOLIZADEH, H. M.; ASSEFA, M. M.; REDDI, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. **sensors**, 2016.
- GHOSH, A.; JOSHI, P. K. A comparison of selected classification algorithms for mappingbamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, n. 1, p. 298–311, 2014.
- GIBBS, J.; WILDE, F. D.; HECKATHORN, H. A. **Chapter A6. Section 6.8. Use of multiparameter instruments for routine field measurements**Reston, VA, 2012. Disponível em: <<https://doi.org/10.3133/twri09A6.8>>
- GOMES, L. C. et al. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, p. 337–350, 2019.
- GUO, H. et al. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. **Journal of Environmental Sciences**, v. 32, p. 90–101, 2015.

- HALABISKY, M. et al. Reconstructing semi-arid wetland surface water dynamics through spectral mixture analysis of a time series of Landsat satellite images (1984-2011). **Remote Sensing of Environment**, v. 177, p. 171–183, 2016.
- HEINONEN, P.; ZIGLIO, G.; VAN DER BEKEN, A. **Hydrological and limnological aspects of lake monitoring**. [s.l.] John Wiley & Sons, 2008.
- HENGL, T. et al. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. **PeerJ**, v. 2018, n. 8, 2018.
- HOSSEINZADEH, A. et al. Application of artificial neural network and multiple linear regression in modeling nutrient recovery in vermicompost under different conditions. **Bioresource Technology**, v. 303, p. 303–311, 2020.
- IBGE. **Instituto Brasileiro de Geografia e Estatística, IBGE Cidades**. Disponível em: <www.cidades.ibge.gov.br>. Acesso em: 6 maio. 2019.
- IBGE. **Instituto Brasileiro de Geografia e Estatística**. Disponível em: <ibge.gov.br>.
- INMET. **Instituto Nacional de Meteorologia**. Disponível em: <<http://www.inmet.gov.br/portal>>. Acesso em: 17 jun. 2020.
- IQBAL, F.; LUCIEER, A.; BARRY, K. Simplified radiometric calibration for UAS-mounted multispectral sensor. **European Journal of Remote Sensing**, v. 51, n. 1, p. 301–313, 2018.
- JENSEN, J. R. **Sensoriamento Remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos/SP, 2009.
- KIA, O. E.; DOCKSTADER, S. L. Photogrammetric method and system for stitching and stabilizing camera images. **U.S. Patent**, v. 183, n. 8, p. 551–565, 2013.
- KIM, H. C. et al. Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. **Marine Pollution Bulletin**, v. 121, n. 1–2, p. 425–434, 2017.
- KIRK, J. T. **Light and Photosynthesis in Aquatic Ecosystems** Nova Iorque, 2011.
- KROGER, S.; LAW, R. J. S. **Sensing the Sea. Trends Biotechnology**, 2005.
- KUHN, M. **caret: Classification and Regression Training**, 2020. Disponível em: <<https://cran.r-project.org/package=caret>>
- KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. [s.l.] Springer, 2013.
- LABOISSIERE, L. A.; FERNANDES, R. A. S.; LAGE, G. G. Maximum and minimum

stock price forecasting of Brazilian power distribution companies based on artificial neural networks. **Applied Soft Computing Journal**, v. 35, p. 66–74, 2015.

LEE RODGERS, J.; NICEWANDER, W. A. Thirteen Ways to Look at the Correlation Coefficient. **The American Statistician**, v. 42, n. 1, p. 59–66, fev. 1988.

LILLESAND, T. M.; KIEFER, R. W. **Remote sensing and image interpretation**. 3. ed. New York: John Wiley & Sons, 1994.

LIM, J.; CHOI, M. Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea. **Environmental Monitoring and Assessment**, v. 187, n. 6, p. 1–17, 2015.

MATTHEWS, M. W. A current review of empirical procedures of remote sensing in Inland and near-coastal transitional waters. **International Journal of Remote Sensing**, v. 32, n. 21, p. 6855–6899, 2011.

MEYER, H. et al. From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: Do we need hyperspectral information? **International Journal of Applied Earth Observation and Geoinformation**, v. 55, p. 21–31, 2017.

MEYER, H. et al. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. **Environmental Modelling and Software**, v. 101, p. 1–9, 2018.

MICASENSE. **MicaSense Camera Image Processing Workflow**. Disponível em: <<https://agisoft.freshdesk.com/support/solutions/articles/31000148780-micasense-rededge-mx-processing-workflow-including-reflectance-calibration-in-agisoft-metashape-pro>>. Acesso em: 2 mar. 2020.

MICASENSE. **MicaSense RedEdge User Manual**.

MORAES, E. C. **Fundamentos de Sensoriamento Remoto**. São José dos Campos/SP: INPE, 2002a.

MORAES, E. C. **Fundamentos de Sensoriamento Remoto** São José dos Campos – SPINPE, , 2002b.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e Metodologias de Aplicação** Viçosa, 2011.

MUÑOZ-ROMERO, S. et al. Informative variable identifier: Expanding interpretability in feature selection. **Pattern Recognition**, v. 98, 2020.

- MURPHY, K. P. **Machine Learning: A Probabilistic Perspective** **Chance Encounters: Probability in Education** London, England The MIT Press, , 2013.
- NAJAFZADEH, M.; ZEINOLABEDINI, M. Prognostication of waste water treatment plant performance using efficient soft computing models: An environmental evaluation. **Measurement: Journal of the International Measurement Confederation**, v. 138, p. 690–701, 2019.
- NICODEMUS, F. E. Reflectance Nomenclature and Directional Reflectance and Emissivity. **Applied Optics**, v. 9, n. 6, p. 1474, 1970.
- NOBREGA, I. V. **Análise espectral de sistemas aquáticos da Amazônia para a Identificação de componentes opticamente ativos**. [s.l.] Instituto Nacional de Pesquisas Espaciais, 2002.
- NOI, P. T.; DEGENER, J.; KAPPAS, M. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. **Remote Sensing**, v. 9, n. 5, 2017.
- NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 3. ed. São Paulo: Blucher, 2008.
- NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 4. ed. São Paulo: Blucher, 2010.
- OLSON, D. L.; DELEN, D. **Advanced data mining techniques** Springer Science & Business Media, , 2008.
- PETROVI, N. V.; MARKOVI, D. P. ASSESSMENT OF WATER QUALITY OF THE PEK RIVER BASED ON PHYSICOCHEMICAL ANALYSIS. **Archives for Technical Sciences**, v. 13, n. 1, p. 59–66, 2015.
- PINTO, E. Geopolítica da água. **Revista de Geopolítica**, p. 19–32, 2017.
- R CORE TEAM. **R: A language and environment for statistical computing, version 3.3. 1**. Vienna, Austria: R Foundation for Statistical Computing, 2020. Disponível em: <<https://www.r-project.org>>
- REUNANEN, J. Overfitting in making comparisons between variable selection methods. **Journal of Machine Learning Research**, v. 3, p. 1371–1382, 2003.
- RITCHIE, J. C.; SCHIEBE, F. R.; MCHENRY, J. R. Remote sensing of suspended sediments in surface waters. **Photogrammetric Engineering and Remote Sensing**, v. 42, n. 12, p. 1539–1545, 1976.

- ROBERT, E. et al. Monitoring water turbidity and surface suspended sediment concentration of the Bagre Reservoir (Burkina Faso) using MODIS and field reflectance data. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p. 243–251, 2016.
- RUDORFF, C. M. et al. Análise derivativa de dados hiperespectrais medidos em nível de campo e orbital para caracterizar a composição de águas opticamente complexas na Amazônia level to characterize the composition of optically complex waters in the. **Acta Amazônica**, v. 37, n. 2, p. 269–280, 2007.
- SÁENZ, N. A.; PAEZ, D. E.; ARANGO, C. Local algorithm for monitoring Total Suspended Sediments in micro-watersheds usin Drones and remote sensing applications. Case study: Teusacá River, la Calera, Colombia. **Remote Sensing and Spatial Information Sciences**, v. 40, n. 1W4, p. 159–165, 26 ago. 2015.
- SANTOS, A. R. **Apostila de Sensoriamento Remoto** Espírito Santo, 2013.
- SESTINI, M. F. **Variáveis Geomorfológicas no estudo de deslizamentos em Caraguatatuba-SP utilizando imagens TM-Landsat e SIG**. São José dos Campos/SP: Instituto Nacional de Pesquisas Espaciais, 1999.
- SEYHAN, E.; DEKKER, A. Applicatin of remote sensing techniques for water quality monitoring. **Hydrobiological Bulletin**, v. 20, p. 41–50, 1986.
- SHARAF EL DIN, E.; ZHANG, Y.; SULIMAN, A. Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework. **International Journal of Remote Sensing**, v. 38, n. 4, p. 1023–1042, 2017.
- SHAWE-TAYLOR, J.; CRISTIANINI, N. **Kernel methods for pattern analysis**. [s.l.] Cambridge university press, 2004.
- SLOGGETT, D. et al. Operational uses of ocean colour data-Perspectives for the OCTOPUS Programme. **Sensors and Environmental Applications of Remote Sensing**, 1995.
- SONG, H. et al. Comparison of mosaicking techniques for airborne images from consumer-grade cameras. **Journal of Applied Remote Sensing**, v. 10, n. 1, p. 16030, 2016.
- STEVENS, A. et al. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. **PLoS ONE**, v. 8, n. 6, 2013.
- SU, T. C.; CHOU, H. T. Application of multispectral sensors carried on unmanned

aerial vehicle (UAV) to trophic state mapping of small reservoirs: A case study of Tain-Pu reservoir in Kinmen, Taiwan. **Remote Sensing**, v. 7, n. 8, p. 10078–10097, 2015.

SU, T.-C. A study of a matching pixel by pixel (MPP) algorithm to establish an empirical model of water quality mapping, as based on unmanned aerial vehicle (UAV) images. **International Journal of Applied Earth Observation and Geoinformation**, v. 58, p. 213–224, jun. 2017.

TUBELIS, A.; NASCIMENTO, F. J. . **Meteorologia Descritiva: Fundamentos e Aplicações Brasileiras**. São Paulo: Nobel, 2000.

TUNDISI, J. G. **Água no século XXI: Enfrentando a escassez**. São Carlos/SP: Editora RiMa, 2003.

UFV et al. **Mapa de solos do Estado de Minas Gerais** Belo Horizonte Fundação Estadual do Meio Ambiente, , 2010.

WILSON, P. C. Water Quality Notes : Water Clarity (Turbidity , Suspended Solids , and Color). **Department of Soil and Water Sciences**, p. 1–8, 2009.

XU, M. et al. Regionally and Locally Adaptive Models for Retrieving Chlorophyll-a Concentration in Inland Waters From Remotely Sensed Multispectral and Hyperspectral Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, v. PP, p. 1–17, 2019.

ZHANG, L. et al. Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. **Materials and Design**, v. 193, p. 108835, 2020.

ZHANG, T.; HUANG, M.; WANG, Z. Estimation of chlorophyll-a Concentration of lakes based on SVM algorithm and Landsat 8 OLI images. **Environmental Science and Pollution Research**, v. 27, n. 13, p. 14977–14990, 2020.

ZHANG, W. **MARS Applications in Geotechnical Engineering Systems**. Pequim, China: Science Press, 2020.

ZHANG, Y. et al. Sensing an intense phytoplankton bloom in the western Taiwan Strait from radiometric measurements on a UAV. **Remote Sensing of Environment**, v. 198, p. 85–94, 2017.

ZHANG, Y.; LI, A.; DENG, B. Data-driven predictive models for chemical durability of oxide glass under different chemical conditions. **npj Materials Degradation**, p. 1–11, 2020.

ZHAO, D. et al. Estimation of water clarity in Taihu Lake and surrounding rivers using Landsat imagery. **Advances in Water Resources**, v. 34, n. 2, p. 165–173, 2011.

APÊNDICE A – COVARIÁVEIS UTILIZADAS PARA IMPLEMENTAÇÃO DOS MODELOS

Razões de bandas e índices utilizados como covariáveis				
Pontual*	50 cm*	100 cm*	200 cm*	400 cm*
PB1	50B1	100B1	200B1	400B1
PB2	50B2	100B2	200B2	400B2
PB3	50B3	100B3	200B3	400B3
PB4	50B4	100B4	200B4	400B4
PB5	50B5	100B5	200B5	400B5
PB1-PB2	50B1-50B2	100B1-100B2	200B1-200B2	400B1-400B2
PB1-PB3	50B1-50B3	100B1-100B3	200B1-200B3	400B1-400B3
PB1-PB4	50B1-50B4	100B1-100B4	200B1-200B4	400B1-400B4
PB1-PB5	50B1-50B5	100B1-100B5	200B1-200B5	400B1-400B5
PB2-PB3	50B2-50B3	100B2-100B3	200B2-200B3	400B2-400B3
PB2-PB4	50B2-50B4	100B2-100B4	200B2-200B4	400B2-400B4
PB2-PB5	50B2-50B5	100B2-100B5	200B2-200B5	400B2-400B5
PB3-PB4	50B3-50B4	100B3-100B4	200B3-200B4	400B3-400B4
PB3-PB5	50B3-50B5	100B3-100B5	200B3-200B5	400B3-400B5
PB4-PB5	50B4-50B5	100B4-100B5	200B4-200B5	400B4-400B5
PB1/PB2	50B1/50B2	100B1/100B2	200B1/200B2	400B1/400B2
PB1/PB3	50B1/50B3	100B1/100B3	200B1/200B3	400B1/400B3
PB1/PB4	50B1/50B4	100B1/100B4	200B1/200B4	400B1/400B4
PB1/PB5	50B1/50B5	100B1/100B5	200B1/200B5	400B1/400B5
PB2/PB3	50B2/50B3	100B2/100B3	200B2/200B3	400B2/400B3
PB2/PB4	50B2/50B4	100B2/100B4	200B2/200B4	400B2/400B4
PB2/PB5	50B2/50B5	100B2/100B5	200B2/200B5	400B2/400B5
PB3/PB4	50B3/50B4	100B3/100B4	200B3/200B4	400B3/400B4
PB3/PB5	50B3/50B5	100B3/100B5	200B3/200B5	400B3/400B5
PB4/PB5	50B4/50B5	100B4/100B5	200B4/200B5	400B4/400B5
PB1+PB2	50B1+50B2	100B1+100B2	200B1+200B2	400B1+400B2

PB1+PB3	50B1+50B3	100B1+100B3	200B1+200B3	400B1+400B3
PB1+PB4	50B1+50B4	100B1+100B4	200B1+200B4	400B1+400B4
PB1+PB5	50B1+50B5	100B1+100B5	200B1+200B5	400B1+400B5
PB2+PB3	50B2+50B3	100B2+100B3	200B2+200B3	400B2+400B3
PB2+PB4	50B2+50B4	100B2+100B4	200B2+200B4	400B2+400B4
PB2+PB5	50B2+50B5	100B2+100B5	200B2+200B5	400B2+400B5
PB3+PB4	50B3+50B4	100B3+100B4	200B3+200B4	400B3+400B4
PB3+PB5	50B3+50B5	100B3+100B5	200B3+200B5	400B3+400B5
PB4+PB5	50B4+50B5	100B4+100B5	200B4+200B5	400B4+400B5
(PB1-PB2)/(PB1+PB2)	(50B1-50B2)/(50B1+50B2)	(100B1-100B2)/(100B1+100B2)	(200B1-200B2)/(200B1+200B2)	(400B1-400B2)/(400B1+400B2)
(PB1-PB2)/(PB1+PB3)	(50B1-50B2)/(50B1+50B3)	(100B1-100B2)/(100B1+100B3)	(200B1-200B2)/(200B1+200B3)	(400B1-400B2)/(400B1+400B3)
(PB1-PB2)/(PB1+PB4)	(50B1-50B2)/(50B1+50B4)	(100B1-100B2)/(100B1+100B4)	(200B1-200B2)/(200B1+200B4)	(400B1-400B2)/(400B1+400B4)
(PB1-PB2)/(PB1+PB5)	(50B1-50B2)/(50B1+50B5)	(100B1-100B2)/(100B1+100B5)	(200B1-200B2)/(200B1+200B5)	(400B1-400B2)/(400B1+400B5)
(PB1-PB2)/(PB2+PB3)	(50B1-50B2)/(50B2+50B3)	(100B1-100B2)/(100B2+100B3)	(200B1-200B2)/(200B2+200B3)	(400B1-400B2)/(400B2+400B3)
(PB1-PB2)/(PB2+PB4)	(50B1-50B2)/(50B2+50B4)	(100B1-100B2)/(100B2+100B4)	(200B1-200B2)/(200B2+200B4)	(400B1-400B2)/(400B2+400B4)
(PB1-PB2)/(PB2+PB5)	(50B1-50B2)/(50B2+50B5)	(100B1-100B2)/(100B2+100B5)	(200B1-200B2)/(200B2+200B5)	(400B1-400B2)/(400B2+400B5)
(PB1-PB2)/(PB3+PB4)	(50B1-50B2)/(50B3+50B4)	(100B1-100B2)/(100B3+100B4)	(200B1-200B2)/(200B3+200B4)	(400B1-400B2)/(400B3+400B4)
(PB1-PB2)/(PB4+PB5)	(50B1-50B2)/(50B4+50B5)	(100B1-100B2)/(100B4+100B5)	(200B1-200B2)/(200B4+200B5)	(400B1-400B2)/(400B4+400B5)
(PB1-PB3)/(PB1+PB4)	(50B1-50B3)/(50B1+50B4)	(100B1-100B3)/(100B1+100B4)	(200B1-200B3)/(200B1+200B4)	(400B1-400B3)/(400B1+400B4)
(PB1-PB3)/(PB1+PB5)	(50B1-50B3)/(50B1+50B5)	(100B1-100B3)/(100B1+100B5)	(200B1-200B3)/(200B1+200B5)	(400B1-400B3)/(400B1+400B5)
(PB1-PB3)/(PB2+PB3)	(50B1-50B3)/(50B2+50B3)	(100B1-100B3)/(100B2+100B3)	(200B1-200B3)/(200B2+200B3)	(400B1-400B3)/(400B2+400B3)
(PB1-PB3)/(PB2+PB4)	(50B1-50B3)/(50B2+50B4)	(100B1-100B3)/(100B2+100B4)	(200B1-200B3)/(200B2+200B4)	(400B1-400B3)/(400B2+400B4)
(PB1-PB3)/(PB2+PB5)	(50B1-50B3)/(50B2+50B5)	(100B1-100B3)/(100B2+100B5)	(200B1-200B3)/(200B2+200B5)	(400B1-400B3)/(400B2+400B5)
(PB1-PB3)/(PB3+PB4)	(50B1-50B3)/(50B3+50B4)	(100B1-100B3)/(100B3+100B4)	(200B1-200B3)/(200B3+200B4)	(400B1-400B3)/(400B3+400B4)
(PB1-PB2)/(PB3+PB5)	(50B1-50B2)/(50B3+50B5)	(100B1-100B2)/(100B3+100B5)	(200B1-200B2)/(200B3+200B5)	(400B1-400B2)/(400B3+400B5)
(PB1-PB4)/(PB1+PB4)	(50B1-50B4)/(50B1+50B4)	(100B1-100B4)/(100B1+100B4)	(200B1-200B4)/(200B1+200B4)	(400B1-400B4)/(400B1+400B4)
(PB1-PB5)/(PB1+PB5)	(50B1-50B5)/(50B1+50B5)	(100B1-100B5)/(100B1+100B5)	(200B1-200B5)/(200B1+200B5)	(400B1-400B5)/(400B1+400B5)
(PB2-PB3)/(PB2+PB3)	(50B2-50B3)/(50B2+50B3)	(100B2-100B3)/(100B2+100B3)	(200B2-200B3)/(200B2+200B3)	(400B2-400B3)/(400B2+400B3)
(PB2-PB4)/(PB2+PB4)	(50B2-50B4)/(50B2+50B4)	(100B2-100B4)/(100B2+100B4)	(200B2-200B4)/(200B2+200B4)	(400B2-400B4)/(400B2+400B4)
(PB3-PB4)/(PB3+PB4)	(50B3-50B4)/(50B3+50B4)	(100B3-100B4)/(100B3+100B4)	(200B3-200B4)/(200B3+200B4)	(400B3-400B4)/(400B3+400B4)
(PB3-PB5)/(PB3+PB5)	(50B3-50B5)/(50B3+50B5)	(100B3-100B5)/(100B3+100B5)	(200B3-200B5)/(200B3+200B5)	(400B3-400B5)/(400B3+400B5)

(PB1-PB5)/(PB2+PB3)	(50B1-50B5)/(50B2+50B3)	(100B1-100B5)/(100B2+100B3)	(200B1-200B5)/(200B2+200B3)	(400B1-400B5)/(400B2+400B3)
(PB1-PB5)/(PB2+PB4)	(50B1-50B5)/(50B2+50B4)	(100B1-100B5)/(100B2+100B4)	(200B1-200B5)/(200B2+200B4)	(400B1-400B5)/(400B2+400B4)
(PB1-PB5)/(PB2+PB5)	(50B1-50B5)/(50B2+50B5)	(100B1-100B5)/(100B2+100B5)	(200B1-200B5)/(200B2+200B5)	(400B1-400B5)/(400B2+400B5)
(PB1-PB5)/(PB3+PB4)	(50B1-50B5)/(50B3+50B4)	(100B1-100B5)/(100B3+100B4)	(200B1-200B5)/(200B3+200B4)	(400B1-400B5)/(400B3+400B4)
(PB1-PB5)/(PB3+PB5)	(50B1-50B5)/(50B3+50B5)	(100B1-100B5)/(100B3+100B5)	(200B1-200B5)/(200B3+200B5)	(400B1-400B5)/(400B3+400B5)
(PB2-PB3)/(PB2+PB4)	(50B2-50B3)/(50B2+50B4)	(100B2-100B3)/(100B2+100B4)	(200B2-200B3)/(200B2+200B4)	(400B2-400B3)/(400B2+400B4)
(PB2-PB3)/(PB2+PB5)	(50B2-50B3)/(50B2+50B5)	(100B2-100B3)/(100B2+100B5)	(200B2-200B3)/(200B2+200B5)	(400B2-400B3)/(400B2+400B5)
(PB2-PB3)/(PB3+PB4)	(50B2-50B3)/(50B3+50B4)	(100B2-100B3)/(100B3+100B4)	(200B2-200B3)/(200B3+200B4)	(400B2-400B3)/(400B3+400B4)
(PB2-PB3)/(PB3+PB5)	(50B2-50B3)/(50B3+50B5)	(100B2-100B3)/(100B3+100B5)	(200B2-200B3)/(200B3+200B5)	(400B2-400B3)/(400B3+400B5)
(PB2-PB3)/(PB4+PB5)	(50B2-50B3)/(50B4+50B5)	(100B2-100B3)/(100B4+100B5)	(200B2-200B3)/(200B4+200B5)	(400B2-400B3)/(400B4+400B5)
(PB2-PB4)/(PB2+PB5)	(50B2-50B4)/(50B2+50B5)	(100B2-100B4)/(100B2+100B5)	(200B2-200B4)/(200B2+200B5)	(400B2-400B4)/(400B2+400B5)
(PB2-PB4)/(PB3+PB4)	(50B2-50B4)/(50B3+50B4)	(100B2-100B4)/(100B3+100B4)	(200B2-200B4)/(200B3+200B4)	(400B2-400B4)/(400B3+400B4)
(PB2-PB4)/(PB3+PB5)	(50B2-50B4)/(50B3+50B5)	(100B2-100B4)/(100B3+100B5)	(200B2-200B4)/(200B3+200B5)	(400B2-400B4)/(400B3+400B5)
(PB2-PB4)/(PB4+PB5)	(50B2-50B4)/(50B4+50B5)	(100B2-100B4)/(100B4+100B5)	(200B2-200B4)/(200B4+200B5)	(400B2-400B4)/(400B4+400B5)
(PB2-PB5)/(PB2+PB5)	(50B2-50B5)/(50B2+50B5)	(100B2-100B5)/(100B2+100B5)	(200B2-200B5)/(200B2+200B5)	(400B2-400B5)/(400B2+400B5)
(PB2-PB5)/(PB3+PB4)	(50B2-50B5)/(50B3+50B4)	(100B2-100B5)/(100B3+100B4)	(200B2-200B5)/(200B3+200B4)	(400B2-400B5)/(400B3+400B4)
(PB2-PB5)/(PB3+PB5)	(50B2-50B5)/(50B3+50B5)	(100B2-100B5)/(100B3+100B5)	(200B2-200B5)/(200B3+200B5)	(400B2-400B5)/(400B3+400B5)
(PB2-PB5)/(PB4+PB5)	(50B2-50B5)/(50B4+50B5)	(100B2-100B5)/(100B4+100B5)	(200B2-200B5)/(200B4+200B5)	(400B2-400B5)/(400B4+400B5)
(PB3-PB4)/(PB3+PB5)	(50B3-50B4)/(50B3+50B5)	(100B3-100B4)/(100B3+100B5)	(200B3-200B4)/(200B3+200B5)	(400B3-400B4)/(400B3+400B5)
(PB3-PB4)/(PB4+PB5)	(50B3-50B4)/(50B4+50B5)	(100B3-100B4)/(100B4+100B5)	(200B3-200B4)/(200B4+200B5)	(400B3-400B4)/(400B4+400B5)
(PB3-PB5)/(PB4+PB5)	(50B3-50B5)/(50B4+50B5)	(100B3-100B5)/(100B4+100B5)	(200B3-200B5)/(200B4+200B5)	(400B3-400B5)/(400B4+400B5)
(PB4-PB5)/(PB4+PB5)	(50B4-50B5)/(50B4+50B5)	(100B4-100B5)/(100B4+100B5)	(200B4-200B5)/(200B4+200B5)	(400B4-400B5)/(400B4+400B5)
Irededge	Irededge.5	Irededge.1	Irededge.2	Irededge.4
NDVIVIS	NDVIVIS.5	NDVIVIS.1	NDVIVIS.2	NDVIVIS.4
GLI	GLI.5	GLI.1	GLI.2	GLI.4
ARVI	ARVI.5	ARVI.1	ARVI.2	ARVI.4
GNDVI	GNDVI.5	GNDVI.1	GNDVI.2	GNDVI.4
NPQI	NPQI.5	NPQI.1	NPQI.2	NPQI.4

*Diferentes tamanhos utilizados para extração das amostras espectrais da imagem

Covariáveis Ambientais, temporais e locais
Batimetria
X
Y
Mês
Lagoa